

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2004, 22:33:30 ; Search time 29.7189 Seconds
(without alignments)
479.034 Million cell updates/sec

Title: US-10-074-225A-5
Perfect score: 966
Sequence: 1 HPHKHSHEQHPGHHPAH.....PSFPLPHKHPLKPDNQPPF.148

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	966	100.0	525	1 KGHUGH	histidine-rich gly
2	503.5	52.1	445	2 A60488	histidine-rich gly
3	241.5	25.0	351	1 KGZQHL	histidine-rich gly
4	217	22.5	140	2 A54523	histidine-rich pro
5	193	20.0	735	2 T45059	hypothetical prote
6	176	18.2	1571	2 T14155	zinc finger protei
7	173.5	18.0	471	2 T33997	hypothetical prote
8	171.5	17.8	657	2 A29454	knob-associated hi
9	170.5	17.7	380	2 H83994	hypothetical prote
10	167	17.3	361	2 F87286	cation efflux fami
11	166.5	17.2	369	2 I78877	CAMP responsive el
12	166.5	17.2	508	2 A45477	CAMP response elem
13	166.5	17.2	515	2 T23089	hypothetical prote
14	166.5	17.2	1378	2 T30173	zinc finger protei
15	162	16.8	436	2 I49714	MHC H-2K/t-w5-link
16	160	16.6	427	2 A32372	female-specific do
17	160	16.6	549	2 B32372	male-specific doub
18	158	16.4	428	2 T27544	zinc resistance pr
19	155.5	16.1	774	2 T39539	alpha-amylase homo
20	155	16.0	199	2 T48099	hypothetical prote
21	149.5	15.5	659	2 A36664	S59/2 homeotic pro
22	149.5	15.5	1002	2 T43236	carboxypeptidase C
23	148.5	15.4	242	2 H82061	hypothetical prote
24	147	15.2	373	2 AD0262	probable membrane
25	147	15.2	375	2 AB3110	conserved hypothet
26	147	15.2	375	2 B98177	hypothetical prote
27	145.5	15.1	654	2 B71623	knob-associated Hi
28	145.5	15.1	1398	2 T13741	hypothetical prote
29	144.5	15.0	60	2 C64698	probable histidine

30 143.5 14.9 852 2 A34373 histidine-rich cal
31 143 14.8 731 1 JC2464 probable copper-tr
32 143 14.8 940 2 JE0291 FB19 protein - hum
33 142.5 14.8 102 2 T30119 hypothetical prote
34 142.5 14.8 270 2 A26480 knob protein - mal
35 142.5 14.8 473 2 A54494 knob-associated hi
36 142.5 14.8 634 2 A54495 knob protein precu
37 142.5 14.8 634 2 A28412 histidine-rich pro
38 142 14.7 301 2 S31782 HRP11 protein - ma
39 142 14.7 997 2 T28872 hypothetical prote
40 141 14.6 905 2 AC2680 hypothetical prote
41 141 14.6 916 2 A97462 hypothetical prote
42 140 14.5 297 2 S23737 proline-rich prote
43 140 14.5 309 2 A25942 histidine/alanine-
44 139 14.4 130 2 S14983 extensin class I (
45 139 14.4 330 2 S22140 nodulin Enod2 - Se

ALIGNMENTS

RESULT 1

KGHUGH

histidine-rich glycoprotein precursor - human

N;Alternate names: HRG

C;Species: Homo sapiens (man)

C;Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 16-Jun-2000

C;Accession: A01287; S29669

R;Koide, T.; Foster, D.; Yoshitake, S.; Davie, E.W.

Biochemistry 25, 2220-2225, 1986

A;Title: Amino acid sequence of human histidine-rich glycoprotein derived from the nucleolus

A;Reference number: A01287; MUID:86216149; PMID:3011081

A;Accession: A01287

A;Molecule type: mRNA

A;Residues: 1-525 <XOI>

A;Cross-references: GB:AB005803; NID:g2280513; PIDN:BAA21613.1; PID:g2280514

R;Hennis, B.; Havelaar, A.; Klufft, C.

submitted to the EMBL Data Library, October 1991

A;Description: PCR detection of a dinucleotide repeat in the human histidine-rich glycoprotein

A;Reference number: S29669

A;Accession: S29669

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 214-247 <HEN>

A;Cross-references: EMBL:Z17218; NID:g32453; PIDN:CAA78925.1; PID:g32454

C;Comment: Although its physiological function is not yet known, HRG does bind heme, dy(din, and the lysine-binding site of plasminogen. On the basis of its homology with HMW) lood coagulation cascade.

C;Comment: The amino half of this protein is homologous to the first two cystatin-like (ould not have inhibitory activity.

C;Comment: In addition to having a high histidine and proline content, this protein has e-rich' region.

C;Genetics:

A;Gene: GDB:HRG

A;Cross-references: GDB:120055; OMIM:142640

A;Map position: 3q27-3q27

C;Superfamily: histidine-rich glycoprotein; cystatin homology

C;Keywords: duplication; glycoprotein; heparin binding; tandem repeat

F;1-18/Domain: signal sequence #status predicted <SIG>

F;19-525/Product: histidine-rich glycoprotein #status predicted <MAT>

F;19-131/Domain: cystatin homology <CY1>

F;140-246/Domain: cystatin homology <CY2>

F;276-321/Region: proline-rich

F;348-437/Region: histidine-rich

F;351-497/Region: proline-rich

F;63,125,344,345/Binding site: carbohydrate (Asn) (covalent) #status predicted

F;78-89,105-126,218-241/Disulfide bonds: #status predicted

Query Match 100.0%; Score 966; DB 1; Length 525;

Best Local Similarity 100.0%; Pred.No. 8.8e-75;

Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HPHKHSHEQHPGHHPAHHPHEHDTHQHPGHHPGHHPGHHPGHHPCH 60

Db 350 HPKHSHSQHPHGHHPHHAHPHEHDTHQPHGHHPHGHHPHGHHPHGHHPHGH 409
QY 61 DFQDYGCDPPPHNQHCCHGHPGPPGHLRRRPGKGRPPFHCRCQIGSVYRLPPLRKGEV 120
Db 410 DFQDYGCDPPPHNQHCCHGHPGPPGHLRRRPGKGRPPFHCRCQIGSVYRLPPLRKGEV 469
QY 121 LPLPEANFPSPPLPHHKHPLKPDNQPPF 148
Db 470 LPLPEANFPSPPLPHHKHPLKPDNQPPF 497

RESULT 2
A60488
histidine-rich glycoprotein - bovine (fragments)
N;Alternate names: autorosette inhibition factor
C;Species: Bos primigenius taurus (cattle)
C;Date: 19-Mar-1993 #sequence_revision 23-Mar-1995 #text_change 07-Jul-1995
C;Accession: S35687; JC2196; A60488
R;Sorensen, C.B.; Krogh-Pedersen, H.; Petersen, T.E.
FEBS Lett. 328, 285-290, 1993
A;Title: Determination of the disulphide bridge arrangement of bovine histidine-rich gly
A;Reference number: S35687; MUID:93351678; PMID:8348977
A;Accession: S35687
A;Molecule type: protein
A;Residues: 1-25;26-52;57-75;82-88;95-119;146-173;175-206;210-309;313-445 <SOR>
A;Note: 355-Gln and 368-Tyr were also found
R;Halkier, T.; Andersen, H.; Vestergaard, A.; Magnusson, S.
Biochem. Biophys. Res. Commun. 200, 78-82, 1994
A;Title: Bovine histidine-rich glycoprotein is a substrate for bovine plasma factor XIII
A;Reference number: JC2196; MUID:94220160; PMID:7909439
A;Accession: JC2196
A;Molecule type: protein
A;Residues: 1-23;35-54,'VK',57-101,'R','TVGEYTEG',116,'N',118,'R',120-136;137-145;150-20
A;Experimental source: plasma
R;Vestergaard, A.B.; Andersen, H.F.; Magnusson, S.; Halkier, T.
Thromb. Res. 60, 385-396, 1990
A;Title: Histidine-rich glycoprotein inhibits contact activation of blood coagulation.
A;Reference number: A60488; MUID:91196010; PMID:2084959
A;Accession: A60488
A;Molecule type: protein
A;Residues: 1-6,'X',8-15 <VES>
C;Comment: This protein is a single-chained plasma protein which participates in transgl
C;Superfamily: histidine-rich glycoprotein; cystatin homology
C;Keywords: glycoprotein; plasma
F;2-113/Domain: cystatin homology <CY1>
F;122-207/Domain: cystatin homology (fragments) <CY2>
F;7-424,60-71,87-108,165-346,180-203,258-288/Disulfide bonds: #status experimental
F;74,107,164,266/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 52.1%; Score 503.5; DB 2; Length 445;
Best Local Similarity 49.2%; Pred. No. 1.3e-35;
Matches 91; Conservative 15; Mismatches 26; Indels 53; Gaps 6;
QY 1 HPKHSHSQHPHGHHPHHAHPHEHDTHQPHGHHPHGHHPHGHHPHGHHPHGH 23
Db 249 HPHESYNFRCPPLLEHKNHSDSPFPQARAPLPPPPGLRCPHPFPFGTKGNRPPPHDSSD 308
QY 24 EHDTHRQHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGH 83
Db 309 EH-----HNHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGH 352
QY 84 PPPGHLRRRPGKGRPPFHCRCQIGSVYRLPPLRKGEVLPLPEANFPSPPLPHHKHPLKPD 143
Db 353 PPPRHSKERGPGKGRPPFHCRCQIGSVYRLPPLRKGEVLPLPEANFPSPPLPHHKHPLKPD 412
QY 144 NQPPF 148
Db 413.IQAPP 417

RESULT 3
KGZQHL

histidine-rich glycoprotein precursor - Plasmodium lophurae
C;Species: Plasmodium lophurae
C;Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 16-Jul-1999
C;Accession: A22692
R;Ravetch, J.V.; Feder, R.; Pavlovic, A.; Blobel, G.
Nature 312, 616-620, 1984
A;Title: Primary structure and genomic organization of the histidine-rich protein of the
A;Reference number: A22692; MUID:85061618; PMID:6095114
A;Accession: A22692
A;Molecule type: DNA
A;Residues: 1-351 <RAV>
A;Cross-references: GB:X01469; NID:g9997; PIDN:CAA25698.1; PID:g9999
C;Comment: There are two copies of 16-residue repeats, two copies of 17-residue repeats,
C;Genetics:
A;Introns: 23/3
C;Superfamily: plasmodium histidine-rich protein
C;Keywords: glycoprotein; tandem repeat
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-47/Domain: propeptide #status predicted <PRO>
F;48-351/Product: histidine-rich glycoprotein #status predicted <MAT>
F;59-74,75-90/Region: 16-residue repeats
F;91-107,108-123/Region: 17-residue repeats
F;124-138,139-153/Region: 15-residue repeats
F;173-301,312-331/Region: 10-residue repeats
F;40/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 25.0%; Score 241.5; DB 1; Length 351;
Best Local Similarity 47.7%; Pred. No. 1.8e-13;
Matches 42; Conservative 1; Mismatches 34; Indels 11; Gaps 3;
QY 2 PKKHSHEQHPHGH-----PHAHPHEH--DTHRQHPHGHHPHGHHPHGHHPHGH 54
Db 184 PHHHHHHHAPHHHHHHHHAPHHHHHHHHAPHHHHHHHHAPHHHHHHHHAPHHHHHH 243
QY 55 HHPHCHDFQDYGCDPPPHNQHCCHGH 82
Db 244 HHHHHHHHD-----AHHHHHHHDAHHH 267

RESULT 4
A54523
histidine-rich protein - Plasmodium lophurae (fragment)
C;Species: Plasmodium lophurae
C;Date: 15-Oct-1994 #sequence_revision 15-Oct-1994 #text_change 20-Aug-1999
C;Accession: A54523
R;Irving, D.O.; Cross, G.A.M.; Feder, R.; Wallach, M.
Mol. Biochem. Parasitol. 18, 223-234, 1986
A;Title: Structure and organization of the histidine-rich protein gene of Plasmodium lop
A;Reference number: A54523; MUID:86174893; PMID:3007981
A;Accession: A54523
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-140 <IRV>
A;Cross-references: GB:M15317; NID:g160331; PIDN:AAA29616.1; PID:g552196
C;Superfamily: plasmodium histidine-rich protein
C;Keywords: tandem repeat
Query Match 22.5%; Score 217; DB 2; Length 140;
Best Local Similarity 39.6%; Pred. No. 9.4e-12;
Matches 38; Conservative 2; Mismatches 34; Indels 22; Gaps 2;
QY 2 PKKHSHEQHPHGHHPHHAHPHEHDTHQPHGHHPHGHHPHGHHPHGHHPHGH 46
Db 8 PHHHHHHHAPHHHHHHHHAPHHHHHHHHAPHHHHHHHHAPHHHHHHHHAPHHHHHH 67
QY 47 PHGHHPHGHHPHCHDFQDYGCDPPPHNQHCCHGH 82
Db 68 HHHHPWFHHHHHHDAHHH-----HHHHHDAHHH 96

RESULT 5
T45059
hypothetical protein Y39B6B.gg [imported] - Caenorhabditis elegans

C;Species: Caenorhabditis elegans
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C;Accession: T45059
R;Wilson, R.; Ainscough, R.; Anderson, K.; Baynes, C.; Berks, M.; Bonfield, J.; Burton, R.; Fulton, L.; Gardner, A.; Green, P.; Hawkins, T.; Hillier, L.; Jier, M.; Johnson, B.; O'Callaghan, M.; Parsons, J.; Percy, C.; Rifkin, L.; Roopra, A.; Saunders, D.; Nature 368, 32-38, 1994
A;Authors: Shownkeen, R.; Sims, M.; Smaldon, N.; Smith, A.; Smith, M.; Sonnhammer, E.; Stoeckli, L.; Wilkinson-Sproat, J.; Wohldman, P.
A;Title: 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans.
A;Reference number: S43531; MUID:94150718; PMID:7906398
A;Accession: T45059
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-735 <WIL>
A;Cross-references: EMBL:AL132896; NID:G6434440; PIDN:CAB60938.1; PID:G6434473
A;Experimental source: clone Y39B6B
C;Genetics:
A;Map position: 3
A;Introns: 18/1; 69/1
A;Note: Y39B6B.gg

Query Match 20.0%; Score 193; DB 2; Length 735;
Best Local Similarity 27.9%; Pred. No. 4.8e-09;
Matches 60; Conservative 4; Mismatches 65; Indels 86; Gaps 11;
QY 1 HPHKHS-HEQHPGHHPHAAHPH-----EHDTHRQH 31
Db 434 HAPAHGHGESHGHHSPAHGHGHEHHHAPAHGHGHEHHHAPAHGHGHEGTHHGH 493
QY 32 PHGHPH--HGHH-----PHGHH-----PHGHH-----PHCHDF 62
Db 494 HGSHPAHGHGHEHHHAPAHGHGHEGTHHGHGHEHHHAPAHGHGHEGTHHGH 553
QY 63 QDYGPC-----DPPHNOGHCHG-HGPPPGHLRRGPG-----KGPRPFHC 103
Db 554 SHSPAHGHGHEHHHAPAHGHGHGSHGVHGHGESHGHGHPAHGHGHEGVHH 613
QY 104 RQIGSVYRLPLRKGEVLPLPEANFFSPLPHKH 138
Db 614 GHGAGYGAHGHGHA-----HHHHAPHHEH 639

RESULT 6
T14155
zinc finger protein Peg3 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T14155
R;Kuroiwa, Y.; Kaneko-Ishino, T.; Kagitani, F.; Kohda, T.; Li, L.L.; Tada, M.; Suzuki, R. Nature Genet. 12, 186-190, 1996
A;Title: Peg3 imprinted gene on proximal chromosome 7 encodes for a zinc finger protein.
A;Reference number: Z17892; MUID:96154192; PMID:8563758
A;Accession: T14155
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1571 <KUR>
A;Cross-references: EMBL:AF038939; NID:G2791677; PID:G2791678; PIDN:AAB96922.1
C;Genetics:
A;Gene: Peg3
A;Map position: 7
C;Keywords: zinc finger

Query Match 18.2%; Score 176; DB 2; Length 1571;
Best Local Similarity 36.5%; Pred. No. 2.7e-07;
Matches 31; Conservative 5; Mismatches 45; Indels 4; Gaps 1;
QY 2 PHKHSHEQHPGHHPHAAHPHEHDTHRQHPGHHPHGHHPHGHHPHGHHPHCHD 61
Db 977 PLVQEMRSEEPHDDKPHGQEPHDDKPHGQEPHDDKPHGQEPHDDKPHGQEPHDDK 1036
QY 62 FQDYGPCDPPPHNQGHCHGHP 86

Db 1037 PIDQEMRSEEPHSE-----ESHGDEP 1057
RESULT 7
T33997
hypothetical protein W03G1.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T33997
R;Pauley, A.; Scheet, P.; Harper, M.
submitted to the EMBL Data Library, February 1999
A;Description: The sequence of C. elegans cosmid W03G1.
A;Reference number: Z21454
A;Accession: T33997
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-471 <PAU>
A;Cross-references: EMBL:AF125964; PIDN:AAD14753.1; GSPDB:GN00022; CESP:W03G1.5
A;Experimental source: strain Bristol N2; clone W03G1
C;Genetics:
A;Gene: CESP:W03G1.5
A;Map position: 4

Query Match 18.0%; Score 173.5; DB 2; Length 471;
Best Local Similarity 40.6%; Pred. No. 1.4e-07;
Matches 39; Conservative 2; Mismatches 36; Indels 19; Gaps 6;
QY 3 HKHSHEQHPGHCH---HPHAHPHEHDTHRQHPGHHPHGHHPH---HGHPHGH 51
Db 369 HGSRSHPRGHGHGCRHGPPHCPGRHG-HHGPPHHHHHDDGRSPSRHGHHHHHGGCRFP 427
QY 52 PHGHPHCHDFQDYGPCDP-----PPHNOGHCHGH 82
Db 428 PHGHHHFPFPF--WPPCPPPPPFWPPHRRGGHCHH 461

RESULT 8
A29454
knob-associated histidine-rich protein precursor - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 09-Jun-2000
C;Accession: A29454
R;Triglia, T.; Stahl, H.D.; Crewther, P.E.; Scanlon, D.; Brown, G.V.; Anders, R.F.; Kemp EMBO J. 6, 1413-1419, 1987
A;Title: The complete sequence of the gene for the knob-associated histidine-rich protein
A;Reference number: A29454; MUID:87275836; PMID:3301325
A;Accession: A29454
A;Molecule type: DNA
A;Residues: 1-657 <TRI>
A;Cross-references: GB:Y00060; NID:G9908; PIDN:CAA68268.1; PID:G9909
C;Superfamily: knob-associated histidine-rich protein

Query Match 17.8%; Score 171.5; DB 2; Length 657;
Best Local Similarity 39.0%; Pred. No. 2.9e-07;
Matches 32; Conservative 5; Mismatches 32; Indels 13; Gaps 2;
QY 7 SHEQHPGHHPHAAHPHEHDTHRQHPGHHPHGHHPHGHHPH---HG-----HHPH 53
Db 57 AQKQHEHHHHHHHHHHHQHQAHPHQAHHHHHHHGEVNHQAQVHQVHQDQAHHHH 116
QY 54 GHHPHCHDFQDYGPCDPPPHNQ 75
Db 117 HHHHLHPQQPQGTVANPPSNE 138

RESULT 9
H83994
hypothetical protein BH2760 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: H83994
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai Nucleic Acids Res. 28, 4317-4331, 2000

9/10/2017

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model
Run on: September 23, 2004, 21:23:24 ; Search time 101.044 Seconds
(without alignments)
413.849 Million cell updates/sec

Title: US-10-074-225A-5
Perfect score: 966
Sequence: 1 HPHKHSHEQHPHGHHPHAH.....PSFPLPHKHPLKPDNQPPF 148

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	966	100.0	148	5	ABB79807 Rabbit hi
2	966	100.0	148	5	ABB79806 Human his
3	966	100.0	525	5	ABB79804 Human his
4	949	98.2	525	8	ADE76897 Human pro
5	543	56.2	526	5	ABB79805 Rabbit hi
6	302	31.3	75	5	ABB07123 HRGP thro
7	254.5	26.3	245	4	ABG28019 Novel hum
8	251.5	26.0	330	4	ABG25331 Novel hum
9	248.5	25.7	406	4	ABG27250 Novel hum
10	242.5	25.1	233	4	ABG19767 Novel hum
11	241.5	25.0	351	2	AAR24393 Sequence
12	240	24.8	491	4	ABG09949 Novel hum
13	230	23.8	183	4	ABG25337 Novel hum
14	225.5	23.3	598	4	ABG23595 Novel hum
15	221	22.9	292	4	AAM18326 Peptide #
16	221	22.9	292	4	ABB37360 Peptide #
17	221	22.9	292	4	AAM30814 Peptide #
18	221	22.9	292	4	ABB32108 Peptide #
19	221	22.9	292	4	ABB22646 Protein #
20	221	22.9	292	4	AAM70490 Human bon
21	221	22.9	292	4	AAM58050 Human bra
22	221	22.9	292	4	ABG52171 Human liv
23	221	22.9	292	4	AAM05933 Peptide #
24	221	22.9	292	5	ABG40129 Human pep
25	219.5	22.7	227	4	ABG14399 Novel hum

26	215	22.3	124	4	ABG12879	Abg12879 Novel hum
27	210	21.7	82	4	AAM13580	Aam13580 Peptide #
28	210	21.7	82	4	ABB32509	Abb32509 Peptide #
29	210	21.7	82	4	AAM25978	Aam25978 Peptide #
30	210	21.7	82	4	ABB27363	Abb27363 Human pep
31	210	21.7	82	4	ABB18015	Abb18015 Protein #
32	210	21.7	82	4	AAM65721	Aam65721 Human bon
33	210	21.7	82	4	AAM53343	Aam53343 Human bra
34	210	21.7	82	4	ABG47362	Abg47362 Human liv
35	210	21.7	82	4	AAM01333	Aam01333 Peptide #
36	210	21.7	82	5	ABG35350	Abg35350 Human pep
37	209.5	21.7	324	4	ABG20509	Abg20509 Novel hum
38	209	21.6	304	4	ABG08412	Abg08412 Novel hum
39	207.5	21.5	79	4	AAM32905	Aau32905 Novel hum
40	204.5	21.2	372	4	ABB67014	Abb67014 Drosophil
41	196.5	20.3	449	4	ABB71938	Abb71938 Drosophil
42	196.5	20.3	449	5	AAM74634	Aau74634 Oestrogen
43	193	20.0	266	4	ABG03494	Abg03494 Novel hum
44	191.5	19.8	156	4	AAM21094	Aam21094 Peptide #
45	191.5	19.8	156	4	ABB43409	Abb43409 Peptide #

ALIGNMENTS

RESULT 1
ABB79807
ID ABB79807 standard; protein; 148 AA.
XX
AC ABB79807;
XX
DT 25-NOV-2002 (first entry)
XX
DE Rabbit histidine proline rich glycoprotein His/Pro-rich domain.
XX
KW Histidine proline rich glycoprotein; HPGP; rabbit; antiangiogenic;
KW cytostatic; antiarteriosclerotic; antiinflammatory; antidiabetic;
KW gynaecological; antiarthritic; antiulcer; osteopathic; antitumour;
KW ophthalmological; nootropic; neuroprotective; antiparkinsonian; chelator.
XX
OS Oryctolagus cuniculus.
XX
PN WO200264621-A2.
XX
PD 22-AUG-2002.
XX
PF 14-FEB-2002; 2002WO-US004336.
XX
PR 14-FEB-2001; 2001US-0268370P.
XX
PA (ATTE-) ATTENJON LLC.
XX
PI Donate F, Harris S, Plunkett ML, Mazar AP;
XX
DR WPI; 2002-666989/71.
DR P-PSDB; ABN84910.
XX
PT New histidine-proline rich glycoprotein (HPRG) polypeptide, useful for
PT diagnosing or treating diseases associated with undesired cell migration,
PT invasion, proliferation, or angiogenesis, e.g. cancer, atherosclerosis.
XX
PS Claim 1; Page 13; 82pp; English.
XX
CC The present sequence is the protein sequence of the histidine-proline-
CC rich (H/P) domain of rabbit anti-angiogenic histidine proline rich
CC glycoprotein (HPRG, see also ABB79805). Claimed anti-angiogenic
CC polypeptides or peptides comprise: the H/P domain (see ABB79806) of human
CC HPRG; an H/P domain of rabbit HPRG; a variant of these that is capable of
CC inhibiting angiogenesis, endothelial cell proliferation or endothelial
CC tube formation in vitro or in vivo; or a pentapeptide having the generic
CC sequence given in ABB79808. Also claimed are: chemically synthesised or
CC recombinantly produced peptide multimers; a diagnostically or
CC therapeutically labeled anti-angiogenic polypeptide, peptide or peptide

CC multimer; a diagnostically useful HPRG-related composition, comprising
CC the diagnostically labeled polypeptide, peptide or peptide multimer and a
CC carrier; an antibody specific for an epitope of HPRG that is present in
CC the H/P domain of human or rabbit HPRG, and which binds to HPRG or to any
CC of the domains in a way which inhibits the anti-angiogenic activity of
CC HPRG or the domain, or an antigen-binding fragment of the antibody; a
CC method for inhibiting cell migration, cell invasion, cell proliferation
CC or angiogenesis, or for inducing apoptosis; a method for treating a
CC subject having a disease or condition associated with undesired cell
CC migration, invasion, proliferation, or angiogenesis; a method of
CC stimulating or inhibiting angiogenesis in a subject; a method of
CC detecting the presence of HPRG or its cleavage product or its peptide in
CC a biological sample; isolated nucleic acids encoding the polypeptide,
CC peptide or peptide multimer; an expression vector; transformed or
CC transfected cells; a method of providing to a cell, tissue or organ an
CC angiogenesis-inhibitory amount of HPRG, an H/P domain of HPRG or its
CC pentapeptide, or the peptide multimer; an affinity ligand useful for
CC binding to, or isolating, an HPRG-binding molecule or cells expressing
CC the binding molecule, comprising the polypeptide, peptide or peptide
CC multimer, immobilised to a solid support or carrier; and a method of
CC isolating HPRG-binding molecule, or isolating or enriching cells
CC expressing HPRG-binding site or receptor, from a complex mixture. The
CC compositions and methods are useful in diagnosing or treating a disease
CC or condition associated with undesired cell migration, invasion,
CC proliferation, or angiogenesis, such as cancer, atherosclerosis, diabetic
CC retinopathy, inflammation, endometriosis, arthritis, peptic ulcers, or
CC fractures. HPRG is especially useful in inhibiting the growth of primary
CC tumours or metastases, and may also be used in treating neurodegenerative
CC diseases like Alzheimer's or Parkinson's disease. The antibodies are
CC stimulators of angiogenesis and are useful for promoting
CC neovascularization in pertinent disease states, and in various
CC immunoassays
XX
SQ Sequence 148 AA;

Query Match 100.0%; Score 966; DB 5; Length 148;
Best Local Similarity 100.0%; Pred. No. 4e-80;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 HPKHHSHEQHPHGHHPHAEHDTHRQHPGHHPHGHHPHGHHPHGHHPHCH 60
DB 1 HPKHHSHEQHPHGHHPHAEHDTHRQHPGHHPHGHHPHGHHPHGHHPHCH 60
QY 61 DFQDYGCDPPPHNQHCCHGCHGPPPPGHLRRRGGKGRPFHCRQICGVYRLPPLRKGEV 120
DB 61 DFQDYGCDPPPHNQHCCHGCHGPPPPGHLRRRGGKGRPFHCRQICGVYRLPPLRKGEV 120
QY 121 LPLPEANFSPPLPHKHPLKPDNQPP 148
DB 121 LPLPEANFSPPLPHKHPLKPDNQPP 148

RESULT 2
ABB79806
ID ABB79806 standard; protein; 148 AA.

XX ABB79806;

XX 25-NOV-2002 (first entry)

DT Human histidine proline rich glycoprotein His/Pro-rich domain.

DE Histidine proline rich glycoprotein; HPRG; human; antiangiogenic;
XX cytoskeletal; antiarteriosclerotic; antiinflammatory; antidiabetic;
KW synaecological; antiarthritic; antiulcer; osteopathic; antitumour;
KW ophthalmological; nootropic; neuroprotective; antiparkinsonian.

XX Homo sapiens.

OS WO200264621-A2.

PN 22-AUG-2002.

XX

PF 14-FEB-2002; 2002WO-US004336.
XX
PR 14-FEB-2001; 2001US-0268370P.
XX
XX (ATTE-) ATTENUON LLC.
XX
XX Donate F, Harris S, Plunkett ML, Mazar AP;
XX
XX WPI; 2002-666989/71.
DR P-PSDB; ABN84910.
XX
XX
PT New histidine-proline rich glycoprotein (HPRG) polypeptide, useful for
PT diagnosing or treating diseases associated with undesired cell migration,
PT invasion, proliferation, or angiogenesis, e.g. cancer, atherosclerosis.
XX
PS Claim 1; Page 13; 82pp; English.

XX The present sequence is the protein sequence of the histidine-proline-
CC rich (H/P) domain of human anti-angiogenic histidine proline rich
CC glycoprotein (HPRG, see also ABB79804). Claimed anti-angiogenic
CC polypeptides or peptides comprise: the H/P domain of human HPRG; an H/P
CC domain of rabbit HPRG (see ABB12345); a variant of these that is capable
CC of inhibiting angiogenesis, endothelial cell proliferation or endothelial
CC tube formation in vitro or in vivo; or a pentapeptide having the generic
CC sequence given in ABB79808. Also claimed are: chemically synthesised or
CC recombinantly produced peptide multimers; a diagnostically or
CC therapeutically labeled anti-angiogenic polypeptide, peptide or peptide
CC multimer; a diagnostically useful HPRG-related composition, comprising
CC the diagnostically labeled polypeptide, peptide or peptide multimer and a
CC carrier; an antibody specific for an epitope of HPRG that is present in
CC the H/P domain of human or rabbit HPRG, and which binds to HPRG or to any
CC of the domains in a way which inhibits the anti-angiogenic activity of
CC HPRG or the domain, or an antigen-binding fragment of the antibody; a
CC method for inhibiting cell migration, cell invasion, cell proliferation
CC or angiogenesis, or for inducing apoptosis; a method for treating a
CC subject having a disease or condition associated with undesired cell
CC migration, invasion, proliferation, or angiogenesis; a method of
CC stimulating or inhibiting angiogenesis in a subject; a method of
CC detecting the presence of HPRG or its cleavage product or its peptide in
CC a biological sample; isolated nucleic acids encoding the polypeptide,
CC peptide or peptide multimer; an expression vector; transformed or
CC transfected cells; a method of providing to a cell, tissue or organ an
CC angiogenesis-inhibitory amount of HPRG, an H/P domain of HPRG or its
CC pentapeptide, or the peptide multimer; an affinity ligand useful for
CC binding to, or isolating, an HPRG-binding molecule or cells expressing
CC the binding molecule, comprising the polypeptide, peptide or peptide
CC multimer, immobilised to a solid support or carrier; and a method of
CC isolating HPRG-binding molecule, or isolating or enriching cells
CC expressing HPRG-binding site or receptor, from a complex mixture. The
CC compositions and methods are useful in diagnosing or treating a disease
CC or condition associated with undesired cell migration, invasion,
CC proliferation, or angiogenesis, such as cancer, atherosclerosis, diabetic
CC retinopathy, inflammation, endometriosis, arthritis, peptic ulcers, or
CC fractures. HPRG is especially useful in inhibiting the growth of primary
CC tumours or metastases, and may also be used in treating neurodegenerative
CC diseases like Alzheimer's or Parkinson's disease. The antibodies are
CC stimulators of angiogenesis and are useful for promoting
CC neovascularization in pertinent disease states, and in various
CC immunoassays

XX Sequence 148 AA;

Query Match 100.0%; Score 966; DB 5; Length 148;
Best Local Similarity 100.0%; Pred. No. 4e-80;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HPKHHSHEQHPHGHHPHAEHDTHRQHPGHHPHGHHPHGHHPHGHHPHCH 60
DB 1 HPKHHSHEQHPHGHHPHAEHDTHRQHPGHHPHGHHPHGHHPHGHHPHCH 60
QY 61 DFQDYGCDPPPHNQHCCHGCHGPPPPGHLRRRGGKGRPFHCRQICGVYRLPPLRKGEV 120
DB 61 DFQDYGCDPPPHNQHCCHGCHGPPPPGHLRRRGGKGRPFHCRQICGVYRLPPLRKGEV 120

QY 121 LPLPEANFSPFPLPHKHKPLKPDNQPPF 148
Db 121 LPLPEANFSPFPLPHKHKPLKPDNQPPF 148

RESULT 3

ABB79804
ID ABB79804 standard; protein; 525 AA.
XX
AC ABB79804;
XX
DT 25-NOV-2002 (first entry)
XX
DE Human histidine proline rich glycoprotein.
XX
KW Histidine proline rich glycoprotein; HPRG; human; antiangiogenic;
KW cytostatic; antiarteriosclerotic; antiinflammatory; antidiabetic;
KW gynaecological; antiarthritic; antiulcer; osteopathic; antitumour;
KW ophthalmological; nootropic; neuroprotective; antiparkinsonian; chelator.
XX
OS Homo sapiens.

FH Key Location/Qualifiers
FT Domain 276..321
FT /note= "Proline-rich domain"
FT Domain 350..497
FT /note= "Histidine-proline-rich domain, region also
FT specifically claimed in Claim 1"

PN WO200264621-A2.
XX
PD 22-AUG-2002.
XX
PF 14-FEB-2002; 2002WO-US004336.
XX
PR 14-FEB-2001; 2001US-0268370P.
XX
PA (ATTE-) ATTENUON LLC.

PI Donate F, Harris S, Plunkett ML, Mazar AP;
XX
XX WPI; 2002-666989/71.
DR P-PSDB; ABN84910.
XX
XX New histidine-proline rich glycoprotein (HPRG) polypeptide, useful for
PT diagnosing or treating diseases associated with undesired cell migration,
PT invasion, proliferation, or angiogenesis, e.g. cancer, atherosclerosis.
XX
PS Claim 1; Page 11; 82pp; English.

XX
CC The present sequence is the protein sequence of human histidine proline
CC rich glycoprotein (HPRG), a proteinaceous metal chelator that can be used
CC to inhibit angiogenesis and treat cancer. Claimed anti-angiogenic
CC polypeptides or peptides comprise: the histidine-proline-rich (H/P)
CC domain (see ABB79806) of human HPRG; an H/P domain of rabbit HPRG (see
CC ABB79807); a variant of these that is capable of inhibiting angiogenesis,
CC endothelial cell proliferation or endothelial tube formation in vitro or
CC in vivo; or a pentapeptide having the generic sequence given in ABB79808.
CC Also claimed are: chemically synthesised or recombinantly produced
CC peptide multimers; a diagnostically or therapeutically labeled anti-
CC angiogenic polypeptide, peptide or peptide multimer; a diagnostically
CC useful HPRG-related composition, comprising the diagnostically labeled
CC polypeptide, peptide or peptide multimer and a carrier; an antibody
CC specific for an epitope of HPRG that is present in the H/P domain of
CC human or rabbit HPRG, and which binds to HPRG or to any of the domains in
CC a way which inhibits the anti-angiogenic activity of HPRG or the domain,
CC or an antigen-binding fragment of the antibody; a method for inhibiting
CC cell migration, cell invasion, cell proliferation or angiogenesis, or for
CC inducing apoptosis; a method for treating a subject having a disease or
CC condition associated with undesired cell migration, invasion,
CC proliferation, or angiogenesis; a method of stimulating or inhibiting
CC angiogenesis in a subject; a method of detecting the presence of HPRG or

CC its cleavage product or its peptide in a biological sample; isolated
CC nucleic acids encoding the polypeptide, peptide or peptide multimer; an
CC expression vector; transformed or transfected cells; a method of
CC providing to a cell, tissue or organ an angiogenesis-inhibitory amount of
CC HPRG, an H/P domain of HPRG or its pentapeptide, or the peptide multimer;
CC an affinity ligand useful for binding to, or isolating, an HPRG-binding
CC molecule or cells expressing the binding molecule, comprising the
CC polypeptide, peptide or peptide multimer, immobilised to a solid support
CC or carrier; and a method of isolating HPRG-binding molecule, or isolating
CC or enriching cells expressing HPRG-binding site or receptor, from a
CC complex mixture. The compositions and methods are useful in diagnosing or
CC treating a disease or condition associated with undesired cell migration,
CC invasion, proliferation, or angiogenesis, such as cancer,
CC atherosclerosis, diabetic retinopathy, inflammation, endometriosis,
CC arthritis, peptic ulcers, or fractures. HPRG is especially useful in
CC inhibiting the growth of primary tumours or metastases, and may also be
CC used in treating neurodegenerative diseases like Alzheimer's disease,
CC Parkinson's disease, and amyotrophic lateral sclerosis. The antibodies
CC are stimulators of angiogenesis and are useful for promoting
CC neovascularization in pertinent disease states, and in various
CC immunoassays

XX Sequence 525 AA;

Query Match 100.0%; Score 966; DB 5; Length 525;
Best Local Similarity 100.0%; Pred. No. 1.4e-79;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HPHKHSHEQHPGHHPHAHPHEHDTHQHPGHHPGHHPGHHPGHHPGHHPCH 60
Db 350 HPHKHSHEQHPGHHPHAHPHEHDTHQHPGHHPGHHPGHHPGHHPGHHPCH 409
QY 61 DFQDYGPCDPPPHNQCHCHGHPGPPGHLRRRGKGPGRPFHCRQIGSVYRLPLEKGEV 120
Db 410 DFQDYGPCDPPPHNQCHCHGHPGPPGHLRRRGKGPGRPFHCRQIGSVYRLPLEKGEV 469
QY 121 LPLPEANFSPFPLPHKHKPLKPDNQPPF 148
Db 470 LPLPEANFSPFPLPHKHKPLKPDNQPPF 497

RESULT 4

ADE76897
ID ADE76897 standard; protein; 525 AA.
XX
AC ADE76897;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human protein expressed in a liver disorder #18.
XX
KW human; liver disorder; hyperlipidaemia; hypertension; type II diabetes;
KW tumour; liver; inflammatory disorder; immune response disorder;
KW high-throughput screening; differential gene expression; gene therapy.
XX
OS Homo sapiens.
XX US2003108871-A1.
PN
XX 12-JUN-2003.
XX
XX 30-JUL-2001; 2001US-00919039.
XX
PR 28-JUL-2000; 2000US-0222113P.
XX
PA (KASE/) KASER M R.
XX
XX Kaser MR;
PI
XX WPI; 2004-031227/03.
DR N-PSDB; ADE76896.
XX

Composition comprising several cDNAs that are differentially expressed in

PT treated human C3A liver cell cultures, useful for treating liver PT disorders.

PS Claim 1; SEQ ID NO 62; 41pp; English.

The invention relates to a composition comprising several cDNAs that are differentially expressed in a liver disorder. The composition is useful for treating liver disorder such as hyperlipidaemia, hypertension, type II diabetes, tumours of the liver and disorders of the inflammatory and immune response. The composition is useful for a high-throughput method of screening several molecules or compounds to identify a ligand which specifically binds a cDNA. A protein encoded by the cDNA is useful for a high-throughput method for using a protein to screen several molecules or compounds to identify at least one ligand which specifically binds the protein which involves combining the protein encoded by the cDNA with several of molecules or compounds under conditions to allow specific binding, and detecting specific binding between the protein and a molecule or compound, therefore identifying a ligand which specifically binds the protein. The composition is useful for detecting and quantifying differential gene expression, can be used in gene therapy, to formulate prognosis and to design a treatment regimen and to monitor the efficacy of treatment. The present sequence represents the amino acid sequence of a protein encoded by a cDNA differentially expressed in a liver disorder.

Sequence 525 AA;

Query Match 98.2%; Score 949; DB 8; Length 525;
Best Local Similarity 98.6%; Pred. No. 5e-78;
Matches 146: Conservative 0: Mismatches 2: Indels

QY 1 HPKXKHSHEQHPGHGHHPPHAHHHPHEHDTHQHPHGHHPPHGHHPPHGHHPPHGHHPPHC 60
 DB 350 HPKXKHSHEQHPGHGHHPPHAHHHPHEHDTHQHPHGHHPPHGHHPPHGHHPPHGHHPPHC 409
 QY 61 DFQDYGPCDPPPHNQGHCCGHGPPPGHLRRRGPGKGRPPFHCQIGSVYRLPPLRKGEV 120
 DB 410 DFQDYGPCDPPPHNQGHCCGHGPPPGHLRRRGPGKGRPPFHCQIGSVYRLPPLRKGEV 469
 QY 121 LPLPEANFFSFPLPHHKHPLKPDNQFPF 148
 DB 470 LPLPEANFFSFPLPHHKHPLKPDIQFPF 497

RESULT 5

ABB79805 standard: protein: 526 AA.

AC ABB79805;

DT 25-NOV-2002 (first entry)

DE Rabbit histidine proline rich glycoprotein.

AA	
KW	Histidine proline rich glycoprotein; HPRG; rabbit; antiangiogenic;
KW	cytostatic; antiarteriosclerotic; antiinflammatory; antidiabetic;
KW	gynaecological; antiarthritic; antiulcer; osteopathic; antitumour;
KW	ophthalmological; nootropic; neuroprotective; antiparkinsonian; chelator.

OS Orvctolagus cuniculus.

Key	Location/Qualifiers
EH	

FT	Domain
FT	/note= "Proline-rich domain"
FT	321..421
FT	Domain
FT	/note= "Histidine-proline-ri-
FT	specifically claimed in Clai

PN WO200264621-A2.

22-AUG-2002

14-FEB-2002: 2002WO-US0043336.

XX
PR 14-FEB-2001; 2001US-0268370P.

PA (ATTE-) ATTENUON LLC.

PI Donate F, Harris S, Plunkett ML, Mazar AP;

DR WPI; 2002-666989/71.
DR P-PSDB: ABN84911.

PT New histidine-proline rich glycoprotein (HPRG) polypeptide, useful for
PT diagnosing or treating diseases associated with undesired cell migration,
PT invasion, proliferation, or angiogenesis, e.g. cancer, atherosclerosis.

PS Claim 1: Page 12-13: 82pp: English:

The present sequence is the protein sequence of rabbit histidine proline rich glycoprotein (HPRG), a proteinaceous chelator that can be used to inhibit angiogenesis and treat cancer. Claimed anti-angiogenic polypeptides or peptides comprise: the histidine-proline-rich (H/P) domain (see ABB79806) of human HPRG; an H/P domain of rabbit HPRG (see ABB79807); a variant of these that is capable of inhibiting angiogenesis, endothelial cell proliferation or endothelial tube formation in vitro or in vivo; or a pentapeptide having the generic sequence given in ABB79808. Also claimed are: chemically synthesised or recombinantly produced peptide multimers; a diagnostically or therapeutically labeled anti-angiogenic polypeptide, peptide or peptide multimer; a diagnostically useful HPRG-related composition, comprising the diagnostically labeled polypeptide, peptide or peptide multimer and a carrier; an antibody specific for an epitope of HPRG that is present in the H/P domain of human or rabbit HPRG, and which binds to HPRG or to any of the domains in a way which inhibits the anti-angiogenic activity of HPRG or the domain, or an antigen-binding fragment of the antibody; a method for inhibiting cell migration, cell invasion, cell proliferation or angiogenesis, or for inducing apoptosis; a method for treating a subject having a disease or condition associated with undesired cell migration, invasion, proliferation, or angiogenesis; a method of stimulating or inhibiting angiogenesis in a subject; a method of detecting the presence of HPRG or its cleavage product or its peptide in a biological sample; isolated nucleic acids encoding the polypeptide, peptide or peptide multimer; an expression vector; transformed or transfected cells; a method of providing to a cell, tissue or organ an angiogenesis-inhibitory amount of HPRG, an H/P domain of HPRG or its pentapeptide, or the peptide multimer; an affinity ligand useful for binding to, or isolating, an HPRG-binding molecule or cells expressing the binding molecule, comprising the polypeptide, peptide or peptide multimer, immobilised to a solid support or carrier; and a method of isolating HPRG-binding molecule, or isolating or enriching cells expressing HPRG-binding site or receptor, from a complex mixture. The compositions and methods are useful in diagnosing or treating a disease or condition associated with undesired cell migration, invasion, proliferation, or angiogenesis, such as cancer, atherosclerosis, diabetic retinopathy, inflammation, endometriosis, arthritis, peptic ulcers, or fractures. HPRG is especially useful in inhibiting the growth of primary tumours or metastases, and may also be used in treating neurodegenerative diseases like Alzheimer's disease, Parkinson's disease and amyotrophic lateral sclerosis. The antibodies are stimulators of angiogenesis and are useful for promoting neovascularization in pertinent disease states, and in various immunoassays

Sequence 526 AA;

Query Match 56.2%; Score 543; DB 5; Length 526;
Best Local Similarity 57.6%; Pred. No. 3.7e-41;
Matches 98: Conservative 8; Mismatches 40; Indels

[illegible]

QY 99 RPFHCRQIGSVYRLPPLRKGEVLPPLPEANFSPFLPHKHPLKPDNQPPF 148
Db 449 FPFHWRRIQSVYQLPPLQKGEVLPPLPEANFPQLLRNHTHPLKPEIQPPF 498

RESULT 6
ABB07123
ID ABB07123 standard; protein; 75 AA.
XX
AC ABB07123;
XX
DT 13-MAR-2002 (first entry)
XX
DE HRGP thrombospondin (TSP)-binding domain fragment (residues 443-517).
XX
KW Thrombospondin; TSP-1; histidine-rich glycoprotein; HRGP; cytostatic;
KW cardiovascular; protein therapy; gene therapy; antisense inhibition;
KW ribozyme inhibition; antibody inhibition; angiogenesis; human; cancer;
KW TSP.
XX
OS Homo sapiens.
XX
PN US2001041670-A1.
XX
PD 15-NOV-2001.
XX
PF 05-DEC-2000; 2000US-00730379.
XX
PR 06-DEC-1999; 99US-0169205P.
XX
PA (SIMA/) SIMANTOV R.
PA (SILV/) SILVERSTEIN R L.
XX
PI Simantov R, Silverstein RL;
XX
DR WPI; 2002-017220/02.
XX
PT Thrombospondin-binding molecules and fragments comprising regions of the
PT TSP-1 binding domains of Histidine-Rich Glycoproteins, useful for
PT treating cancers and cardiovascular diseases.
XX
PS Example; Fig 1; 22pp; English.
XX
CC The invention provides thrombospondin (TSP)-binding molecules and
CC fragments comprising regions of the TSP-1 binding domains of Histidine-
CC Rich Glycoproteins (HRGP). The proteins are used for the regulation of
CC cellular growth and proliferation, such as in the accretion of new blood
CC vessels (angiogenesis), particularly for treatment of cardiovascular
CC disease. They may be used for the reduction of angiogenesis, including
CC growth suppression and arrest, and apoptosis in normal development, for
CC example in embryogenesis, and in a wide range of disorders and diseases,
CC including those involving tumours, malignancies, neoplastic and other
CC pathological conditions and homeostatic imbalances in the control of
CC growth and development, in particular anal cancer, bladder cancer, small
CC cell lung cancer, non-small cell lung cancer, bone cancer, brain cancer,
CC breast cancer, cervical cancer, chondrosarcoma, clear cell adenocarcinoma
CC (DES), colorectal cancer, endometrial cancer, kidney cancer, cancer of the
CC of the eye, eyelid, Kaposi's sarcoma, leukemia, liver cancer, lymphoma,
CC melanoma, mesothelioma, oral cancer, ovarian cancer, pancreatic cancer,
CC prostate cancer, skin cancer, squamous cell cancer, stomach cancer,
CC testicular cancer, thyroid cancer, hepatoma, neuroendocrine cancer,
CC liposarcoma, head and neck cancer and a cholangiocarcinoma. The present
CC sequence represents the TSP-binding motif of HRGP
XX
SQ Sequence 75 AA;

Query Match 31.3%; Score 302; DB 5; Length 75;
Best Local Similarity 100.0%; Pred. No. 4.1e-20;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 GPRPFHCRQIGSVYRLPPLRKGEVLPPLPEANFSPFLPHKHPLKPDNQPPF 148

Db 1 GPRPFHCRQIGSVYRLPPLRKGEVLPPLPEANFSPFLPHKHPLKPDNQPPF 52

RESULT 7
ABG28019
ID ABG28019 standard; protein; 245 AA.
XX
AC ABG28019;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #28010.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS92206.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 58378; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activities in
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 245 AA;

Query Match 26.3%; Score 254.5; DB 4; Length 245;
Best Local Similarity 48.2%; Pred. No. 2.8e-15;
Matches 41; Conservative 2; Mismatches 39; Indels 3; Gaps 1;

QY 1 HPHKHSHEQHPGHHPAHHPHEHDTHQHPHGHHPHGHHPHGHHPGHCH--HP 57
Db 25 HYHHHHHHHHYHRRHHHPHPHPHPHYHTHHHHHHHHHHYHQHPHHHHHHHHHH 84

CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological actions. The
CC polypeptide and polynucleotide sequences have application in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences. The invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 183 AA;

Query Match 23.8%; Score 230; DB 4; Length 183;
Best Local Similarity 46.3%; Pred. No. 3.4e-13;
Matches 38; Conservative 1; Mismatches 29; Indels 14; Gaps 1;

QY 1 HPHKHSHEQHPHGHHPHAEHPHEHDTKQHPHGHHPHGHHPHGHHPHGHHPHGH 60
Db 57 HHH 114
QY 61 DFQDYGPCDPPPHNQHCCHGH 82
Db 115 -----HDQHHHHHHH 124

RESULT 14
ABG23595
ID ABG23595 standard; protein; 598 AA.

XX AC ABG23595;

DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #23586.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS87782.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.

XX PS Claim 20; SEQ ID NO 53954; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
sequences. (I) is useful as hybridisation probes, polymerase chain

CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological actions. The
CC polypeptide and polynucleotide sequences have application in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences. The invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 598 AA;

Query Match 23.3%; Score 225.5; DB 4; Length 598;
Best Local Similarity 45.3%; Pred. No. 2.9e-12;
Matches 39; Conservative 4; Mismatches 32; Indels 11; Gaps 3;

QY 1 HPHKHSHEQHPHGH--HPEHHPHEHDTKQHPHGH-HPHGHHPHGHHPHGHHPHGH 56
Db 331 HHH 390
QY 57 PHCHDFQDYGPCDPPPHNQHCCHGH 82
Db 391 YHCHHHHDY-----HHPHCHHHH 409

RESULT 15
AAM18326

ID AAM18326 standard; protein; 292 AA.

XX AC AAM18326;

DT 12-OCT-2001 (first entry)

XX DE Peptide #4760 encoded by probe for measuring cervical gene expression.

XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX KW cervical cancer.

XX OS Homo sapiens.

XX PN WO200157278-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US000670.

XX PR 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-488901/53.

XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
gene expression in human cervical epithelial cells.

```
PS Claim 27; SEQ ID NO 23152; 487pp; English.
```

```
XX
```

```
CC The present invention relates to human single exon nucleic acid probes
```

```
CC (SENP: see AA110068-AA128459). The present sequence is a peptide encoded
```

```
CC by one such probe. The SENPs are derived from human Hela cells. The SENPs
```

```
CC can be used to produce a single exon microarray, which can be used for
```

```
CC measuring human gene expression in a sample derived from human cervical
```

```
CC epithelial cells. By measuring gene expression, the probes are therefore
```

```
CC useful in grading and/or staging of diseases of the cervix, notably
```

```
CC cervical cancer. Note: The sequence data for this patent did not form
```

```
CC part of the printed specification, but was obtained in electronic format
```

```
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
```

```
XX
```

```
SQ Sequence 292 AA;
```

```
Query Match          22.9%; Score 221; DB 4; Length 292;
```

```
Best Local Similarity 44.1%; Pred. No. 3.6e-12;
```

```
Matches 41; Conservative 3; Mismatches 31; Indels 18; Gaps 4;
```

```
QY 3 HKHSHEQHPGHHPPHAAHPDHEHDTHRQHPHGHPGHHPGHGH -HPHGHPGHHPGHCHD 61  
|||: ||| ||| ||| : ||| ||| ||| ||| ||| ||| |||
```

```
Dd 26 HHHHHHDYDH -HHHHHHHYHHHHHHYHHHPHHHHNHHHHHHHHYHYHHHHHHHHHH - 83  
|||: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
QY 62 FDYGPDCPPPHNQHCCHGCGPPPGHLRRRG 94  
||: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Dd 84 -----PHHNHHHHHH-----HHRRHY 101  
||: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: September 23, 2004, 22:50:02
Job time : 105.044 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2004, 21:27:50 ; Search time 17.2369 Seconds
(without alignments)
447.085 Million cell updates/sec

Title: US-10-074-225A-5
Perfect score: 966
Sequence: 1 HPHKHSHSEQPHGHHPAH.....PSFPLPHHKHPLKPDNQPPF 148

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	966	100.0	525	1 HRG HUMAN	P04196 homo sapien
2	543	56.2	526	1 HRG_RABIT	Q28640 oryctolagus
3	489	50.6	396	1 HRG_BOVIN	P33433 bos taurus
4	241.5	25.0	351	1 HRPX_PLALO	P04929 plasmodium
5	196.5	20.3	449	1 CSUP_DROME	Q9v3a4 drosophila
6	171.5	17.8	657	1 KNOB_PLAFN	P06719 plasmodium
7	166.5	17.2	515	1 KE4L_CAEEL	Q9xtq7 caenorhabdi
8	162	16.8	476	1 KE4_MOUSE	Q31125 mus musculu
9	160	16.6	549	1 DSX_DROME	P23023 drosophila
10	155.5	16.1	774	1 AMY2_SCHPO	O42918 schizosacch
11	154	15.9	352	1 KE4_BRARE	Q9pub8 brachydanio
12	154	15.9	428	1 FXB2_MOUSE	Q64733 mus musculu
13	149.5	15.5	659	1 HMN1_DROME	P22807 drosophila
14	149.5	15.5	815	1 PYGO_DROME	Q13849 schizosacch
15	149.5	15.5	1002	1 CBPY_SCHPO	Q92504 homo sapien
16	147.5	15.3	469	1 KE4_HUMAN	Q48251 helicobacte
17	144.5	15.0	59	1 HFN_HELPY	P16230 oryctolagus
18	143.5	14.9	852	1 SRCH_RABIT	P13817 plasmodium
19	142.5	14.8	473	1 KNOB_PLAFA	P09346 plasmodium
20	142.5	14.8	634	1 KNOB_PLAFG	Q9wv48 rattus norv
21	142	14.7	2167	1 SHK1_RAT	P09346 plasmodium
22	141	14.6	402	1 SELP_BOVIN	Q9wv48 rattus norv
23	140	14.5	332	1 HRP1_PLAFA	P05227 plasmodium
24	138.5	14.3	2161	1 SHK1_HUMAN	Q9v566 homo sapien
25	137	14.2	367	1 SEPA_BRARE	Q98sv1 brachydanio
26	137	14.2	419	1 GSC_DROME	P54366 drosophila
27	137	14.2	620	1 EXTN_TOBAC	P13983 nicotiana t
28	136.5	14.1	302	1 HYPB_BRAJA	Q45257 bradyrhizob
29	136.5	14.1	661	1 KNG_MOUSE	O08677 mus musculu
30	136.5	14.1	1319	1 MN1_HUMAN	Q10571 homo sapien
31	133.5	13.8	335	1 HYPB_RHOCA	P26410 rhodobacter
32	133.5	13.8	469	1 FXGA_HUMAN	P55316 homo sapien
33	133	13.8	355	1 OTX1_MOUSE	P80205 mus musculu

34	133	13.8	355	1 OTX1_RAT	Q63410 rattus norv
35	132	13.7	309	1 NO75_SOYBN	P08297 glycine max
36	132	13.7	337	1 ZNUA_HAEIN	P44526 haemophilus
37	132	13.7	395	1 SRY_MOUSE	Q05738 mus musculu
38	131	13.6	653	1 SF01_MOUSE	Q64213 mus musculu
39	130.5	13.5	312	1 EX24_ARATH	Q9f176 arabidopsis
40	130.5	13.5	645	1 BRH2_DROME	Q24256 drosophila
41	129	13.4	354	1 OTX1_HUMAN	P32242 homo sapien
42	129	13.4	622	1 YA41_HUMAN	Q9upw0 homo sapien
43	129	13.4	639	1 SF01_HUMAN	Q15637 homo sapien
44	128.5	13.3	355	1 SRY_MUSSP	Q62563 mus spretus
45	128	13.3	112	1 NO75_PEA	P16329 pisum sativ

ALIGNMENTS

RESULT 1

ID	HRG_HUMAN	STANDARD;	PRT;	525 AA.
AC	P04196;			
DT	20-MAR-1987 (Rel. 04, Created)			
DT	20-MAR-1987 (Rel. 04, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Histidine-rich glycoprotein precursor (Histidine-proline rich glycoprotein) (HPRG).			
GN	HRG.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Liver;			
RX	MEDLINE=86216149; PubMed=3011081;			
RA	Koide T., Foster D.C., Yoshitake S., Davie E.W.;			
RT	"Amino acid sequence of human histidine-rich glycoprotein derived from the nucleotide sequence of its cDNA.";			
RL	Biochemistry 25:2220-2225(1986).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Wakabayashi S., Takahashi K., Tokunaga F., Koide T.;			
RL	Submitted (JUL-1997) to the EMBL/GenBank/DBSJ databases.			
RN	[3]			
RP	SEQUENCE OF 214-247 FROM N.A.			
RX	MEDLINE=94245171; PubMed=8188234;			
RA	Hennis B.C., Frants R.R., Bakker E., Vossen R.H., van der Poort E.W.,			
RA	Blonden L.A., Cox S., Khan P.M., Spurr N.K., Kluft C.;			
RT	"Evidence for the absence of intron H of the histidine-rich glycoprotein (HRG) gene: genetic mapping and in situ localization of HRG to chromosome 3q28-q29.";			
RL	Genomics 19:195-197(1994).			
RN	[4]			
RP	SEQUENCE OF 19-27.			
RC	TISSUE=Plasma;			
RX	MEDLINE=93092937; PubMed=1459097;			
RA	Hughes G.J., Frutiger S., Paquet N., Ravier F., Pasquali C.,			
RA	Sanchez J.-C., James R., Tissot J.-D., Bjellqvist B.,			
RA	Hochstrasser D.F.;			
RT	"Plasma protein map: an update by microsequencing.";			
RL	Electrophoresis 13:707-714(1992).			
CC	-!- FUNCTION: The physiological function is not yet known. It binds heme, dyes and divalent metal ions. It can inhibit rosette formation and is known to interact with heparin, thrombospondin, and the lysine-binding site of plasminogen. On the basis of its homology with HMW kininogen, the His-rich region of this protein may mediate the contact activation phase of intrinsic blood coagulation cascade.			
CC	-!- SUBCELLULAR LOCATION: Secreted.			
CC	-!- TISSUE SPECIFICITY: Expressed by the liver and secreted in plasma.			
CC	-!- DOMAIN: In addition to having a high His and Pro content, this protein has many internal repeats. 12 tandem repetitions of a 5-residue sequence (GHHPH, consensus) form a histidine-rich region.			


```

Db      259 HSSDE-----HHPHGHHPHGHHHPGHHPPDNDFYDHGPCDPPEHR-----P 304
QY      85 PPGLRRRGPKGRPFHCRQIGSVVRLPPLRKGEVLPLPEANFSPFLPHHKHPLKPDN 144
       ||| : ||||| ||| : ||| : ||||| ||||| ||| : ||| : ||| : ||| :
Db     305 PPRHSKERGPCKGHRFRFWRTGYIHRLPSLKKGEVLPLPEANFSPFSLPNNHNPLQPEI 364
       ||| : ||||| ||| : ||| : ||||| ||||| ||| : ||| : ||| : ||| :
QY     145 QFPF 148
       |   ||
Db     365 QAFF 368

RESULT 4
HRPX PLALO
ID`-HRPX PLALO          STANDARD;          PRT;          351 AA.
AC  P04929;
DT  13-AUG-1987 (Rel. 05, Created)
DT  13-AUG-1987 (Rel. 05, Last sequence update)
DT  15-JUL-1999 (Rel. 38, Last annotation update)
DE  Histidine-rich glycoprotein precursor.
OS  Plasmodium lophurae.
OC  Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX  NCBI_TaxID=5853;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=85061618; PubMed=6095114;
RA  Ravetch J.V., Feder R., Pavlovac A., Blobel G.;
RT  "Primary structure and genomic organization of the histidine-rich
    protein of the malaria parasite Plasmodium lophurae.";
RL  Nature 312:616-620(1984).
CC  -!- MISCELLANEOUS: In the intraerythrocytic stages of development of
    P.lophurae in ducks, there is a synthesis of a major protein that
    accumulates to comprise at least 50% of the cellular mass: the
    histidine rich protein.
-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
    between the Swiss Institute of Bioinformatics and the EMBL outstation
    the European Bioinformatics Institute. There are no restrictions on its
    use by non-profit institutions as long as its content is in no way
    modified and this statement is not removed. Usage by and for commercial
    entities requires a license agreement (See http://www.isb-sib.ch/announce/
    or send an email to license@isb-sib.ch).
-----
DR  EMBL; X01469; CAA25698.1; --
DR  PIR; A22692; KGZQHL.
KW  Malaria; Repeat; Glycoprotein; Signal.
FT  SIGNAL              1         23
FT  PROPEP             24        47
FT  CHAIN              48       351
FT  CARBOHYD           40        40
FT  DOMAIN            59        90
FT  REPEAT            59       74
FT  REPEAT            75        90
FT  DOMAIN            91       123
FT  REPEAT            91       107
FT  REPEAT           108       123
FT  DOMAIN           124       153
FT  REPEAT           124       138
FT  REPEAT           139       153
FT  DOMAIN           173       351
SQ  SEQUENCE 351 AA; 44032 MW; D19A48D47D890453 CRC64;
Query Match                25.0%; Score 241.5; DB 1; Length 351;
Best Local Similarity      47.7%; Pred.No.1.1e-12;
Matches 42; Conservative 1; Mismatches 34; Indels 11; Gaps 3;

QY      2 PHKHSHHQHPHGHH-----PHAHPHEH--DTHRQHFGHHHPGHHHPGHHHPGHHHPG 54
       ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     184 PHHHHHHHAPHHHHHHHHHHAPHHHHHHHHHHAPHHHHHHHHHHGHHHHHHHHHHG 243
       ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      55 HHPHCHDFQDYGCDDPPPHNQHCCHG 82
       ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     244 HHHHHHHHD-----AHHHHHHHHDAAHH 267

```



```

DR EMBL; M25293; AAA17841.1; -.
DR EMBL; M25294; AAA17842.1; -.
DR EMBL; AE003676; AAF54168.1; -.
DR PIR; A32372; A32372.
DR PIR; B32372; B32372.
DR PDB; 1LPV; 02-OCT-02.
DR TRANSFAC; T00955; -.
DR TRANSFAC; T00956; -.
DR FlyBase; FBgn0000504; dsx.
DR GO; GO:0003729; F:RNA binding; NAS.
DR GO; GO:0003700; P:transcription factor activity; NAS.
DR GO; GO:0007619; P:courtship behavior; NAS.
DR GO; GO:0045497; P:female anialia morphogenesis (sensu Holometabola); NAS.
DR GO; GO:0007486; P:female genital morphogenesis (sensu Holometabola); NAS.
DR GO; GO:0019101; P:female somatic sex determination; NAS.
DR GO; GO:0007483; P:genital disc metamorphosis; NAS.
DR GO; GO:0045496; P:male anialia morphogenesis (sensu Holometabola); NAS.
DR GO; GO:0045433; P:male courtship behavior (sensu Insecta); so. .; NAS.
DR GO; GO:0007485; P:male genital morphogenesis (sensu Holometabola); NAS.
DR GO; GO:0019102; P:male somatic sex determination; NAS.
DR GO; GO:0045892; P:negative regulation of transcription; NAS.
DR GO; GO:0045893; P:positive regulation of transcription; NAS.
DR GO; GO:0007548; P:sex differentiation; NAS.
DR InterPro; IPR001275; DM_DNA-binding.
DR SMART; SM00751; DM-domain; 1.
DR SMART; SM00301; DM; 1.
DR PROSITE; PS40000; DM_DOMAIN_1; 1.
DR PROSITE; PS0809; DM_DOMAIN_2; 1.
KW Sexual differentiation; Alternative splicing; DNA-binding;
KW Transcription regulation; Nuclear protein; Zinc; Metal-binding;
KW 3D-structure.
FT DNA_BIND 44 91 DM.
FT DOMAIN 119 224 HIS-RICH.
FT DOMAIN 267 296 SER/GLY-RICH.
FT VARSPLIC 398 427 ARVEINRTVAQIYNYITPMALVNGAPMYL -> GQYVYNE
FT YSRQNLNIYDGGELRNTTRQCG (in isoform
FT Female).
FT FTID=VSP 001321.
FT Missing (in isoform Female).
FT /FTID=VSP 001322.
FT VARSPLIC 428 549
FT MUTAGEN 47 47 C->A,H: ABOLISHES DNA-BINDING.
FT MUTAGEN 50 50 H->Y: ABOLISHES DNA-BINDING.
FT MUTAGEN 59 59 H->Y: ABOLISHES DNA-BINDING.
FT MUTAGEN 68 68 C->D,Y: ABOLISHES DNA-BINDING.
FT MUTAGEN 70 70 C->Y: ABOLISHES DNA-BINDING.
FT MUTAGEN 91 91 R->Q: ABOLISHES DNA-BINDING.
SQ SEQUENCE 549 AA; 57409 MW; 3C1B92724E4CE083 CRC64;

Query Match 16.6%; Score 160; DB 1; Length 549;
Best Local Similarity 28.8%; Pred. No. 5.2e-06;
Matches 42; Conservative 3; Mismatches 37; Indels 64; Gaps 9;

QY 1 HPHKHSHEQHPGHHPHHPHE-HDTHRQ-----H 31
Db 128 HVHAHVHVAHHAHGGH-HSHGHVLH--HQQAAAAAAPSAPASHLGGSSSTAASSIGH 184
QY 32 PHGHHPH-----GHHPHGHHPGHHPH--GHHPHCHDFQDYGPCDP-----PP 72
Db 185 AHAHVHMAAAAAASVAQHQQHSHHHHHQHHPHQ-----QPATQTLRSP 237
QY 73 HNOGHCCHGHPGPPGHLRRRPGKGP 98
Db 238 HSD-----HGGVGPATSSSGGAP 257

RESULT 10
AMY2_SCHPO
ID AMY2_SCHPO STANDARD; PRT; 774 AA.
AC O42918; Q96WR2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable alpha-amylase meu7 precursor (EC 3.2.1.1) (1,4-alpha-D-
```

```

DE DE glucan glucanohydrolase) (Meiotic expression upregulated protein 7).
GN MEU7 OR SPBC16A3.13.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders R., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Walsh S.V., Warren T., Whitehead S.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Wambutt R., Purnelle B.,
RA Eger P., Zimmermann W., Wedler H., Gloux S., Lelaure V., Mottier S.,
RA Goffeau A., Cadieu E., Dreano S., Hunt C., Moore K., Hurst S.M.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shipakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
RN [2]
RP SEQUENCE OF 525-774 FROM N.A.
RC STRAIN=CD16-1;
RX MEDLINE=21270454; PubMed=11376151;
RA Watanabe T., Miyashita K., Saito T.T., Yoneki T., Kakiwara Y.,
RA Nabeshima K., Kishi Y.A., Shimoda C., Nojima H.;
RT "Comprehensive isolation of meiosis-specific genes identifies novel
RT proteins and unusual non-coding transcripts in Schizosaccharomyces
RT pombe.";
RL Nucleic Acids Res. 29:2327-2337(2001).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
CC linkages in oligosaccharides and polysaccharides.
CC -!- COFACTOR: Binds 1 calcium ion per subunit (potential).
CC -!- SIMILARITY: Belongs to family 13 of glycosyl hydrolases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL021748; CAAL6864.1; -.
DR EMBL; AB054318; BAB60884.1; -.
DR PIR; T39539; T39539.
DR HSSP; P56271; 2AAA.
DR GeneDB SPombe; SPBC16A3.13; -.
DR InterPro; IPR006047; Alpha_amyl_cat.
DR Pfam; PF00128; alpha-amylase; 1.
KW Meiosis; Carbohydrate metabolism; Hydrolase; Glycosidase;
KW Calcium-binding; Signal; Glycoprotein.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 774 PROBABLE ALPHA-AMYLASE MEU7.
FT ACT_SITE 461 461 BY SIMILARITY.
FT ACT_SITE 553 553 BY SIMILARITY.
```


FT METAL 143 143 CALCIUM 1 (POTENTIAL).
FT METAL 395 395 CALCIUM 1 (VIA CARBONYL OXYGEN)
FT (POTENTIAL).
FT DISULFID 52 60 BY SIMILARITY.
FT DISULFID 229 314 BY SIMILARITY.
FT DISULFID 495 539 BY SIMILARITY.
FT DISULFID 690 724 BY SIMILARITY.
FT CARBOHYD 143 143 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 214 214 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 243 243 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 272 272 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 326 326 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 355 355 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 409 409 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 774 AA; 89459 MW; 92CEAAE45664269B CRC64;

Query Match 16.1%; Score 155.5; DB 1; Length 774;
Best Local Similarity 26.8%; Pred. No. 1.6e-05;
Matches 53; Conservative 12; Mismatches 52; Indels 81; Gaps 14;

QY 3 HKHSHEQHPHGH-----PH-----AHPHEHDTTHRQ-----PHG-----HHPH 38
DB 249 HKPWKEHCCHDKFPRPVPHNGTKPDHKPKWKEEHCCHGKFFRPIPHNGTKPDHKPW 308
QY 39 GHHPHGH-----PHG-----HHPHGHHPHC-----HD-FQDYGCDPPPHN----- 74
DB 309 KHEEHCCHGRFPRPVPHNGTKPDHKPKWKEEHCCHDKFS-----RPVPHNGTKPDHKP 363
QY 75 ---QGHCHGH--GPPPGHLRRRGPGKGRPFHCRQIGSVYRLPLRKGEVLPPLPEANFP 129
DB 364 WXHEEHCCHGKFLRPVPHNVTKPDHKPKWKEEHCCH-----HGKEP 403
QY 130 SFPLPHKHPLKPDNQPF 147
DB 404 R-PVPH--NGTKPDHKPW 418

RESULT 11
KE4 BRARE
ID KE4 BRARE STANDARD; PRT; 352 AA.
AC Q9PUB8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Zinc transporter SLC39A7 homolog (Histidine-rich membrane protein Ke4
DE homolog) (Fragment).
GN HKE4 OR KE4.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Murray B.W., Suelmann H., Klein J.;
RT "Identification of a homolog of the human HKE4 gene in zebrafish."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- SIMILARITY: Belongs to the ZIP transporter (TC 2.A.5) family.
CC KE4/Catsup subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF196345; AAF05821.1; --

DR ZFIN; ZDB-GENE-991110-20; ke4.
DR InterPro; IPR003689; Zn_transpt_zip.
DR Pfam; PF02535; Zip; 1.
KW Transport; Transmembrane; Glycoprotein.
FT TRANSMEM 3 23 POTENTIAL.
FT TRANSMEM 128 148 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
FT TRANSMEM 215 235 POTENTIAL.
FT TRANSMEM 318 338 POTENTIAL.
FT DOMAIN 24 105 HIS-RICH.
FT DOMAIN 177 217 HIS-RICH.
FT CARBOHYD 311 311 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 352 352
SQ SEQUENCE 352 AA; 37922 MW; C8C8C60F6D2BA8A6 CRC64;

Query Match 15.9%; Score 154; DB 1; Length 352;
Best Local Similarity 42.7%; Pred. No. 1e-05;
Matches 35; Conservative 2; Mismatches 23; Indels 22; Gaps 8;

QY 1 HPHKHSH-----EQHPHG-----HH-----PHAHHPHEHDTTHRQHPHG-HPHGHHPHG 44
DB 24 HSHHHGHGDDGGCHGHGAKMHHGASKWSAEANLPHAESEHHVHDHGHGTHDHA-HDHG 82
QY 45 H-HPHGH-HPHGH--HHPCHD 61
DB 83 HAHSGDIDHGHGHAHKGHAHD 104

RESULT 12
FXB2 MOUSE
ID FXB2 MOUSE STANDARD; PRT; 428 AA.
AC Q64733;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Forkhead box protein B2 (Transcription factor FKH-4).
GN FOXB2 OR FKH4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97014266; PubMed=8861101;
RA Kaestner K.H., Schuetz G., Monaghan A.P.;
RT "Expression of the winged helix genes fkh-4 and fkh-5 defines domains
RT in the central nervous system."
RL Mech. Dev. 55:221-230(1996).
RN [2]
RP SEQUENCE OF 4-114 FROM N.A.
RC STRAIN=129;
RX MEDLINE=93361500; PubMed=7689224;
RA Kaestner K.H., Lee K.H., Schloendorff J., Hiemisch H.,
RA Monaghan A.P., Schuetz G.;
RT "Six members of the mouse forkhead gene family are developmentally
RT regulated."
RL Proc. Natl. Acad. Sci. U.S.A. 90:7628-7631(1993).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DEVELOPMENTAL STAGE: Expressed during embryogenesis.
CC -!- SIMILARITY: Contains 1 fork-head domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X92591; CAA63335.1; --
CC EMBL; X71942; CAA50744.1; --
CC PIR; D47746; D47746.
CC HSSP; Q63245; 2HFH.

DR TRANSFAC; T02442; --
DR MGD; MGI:1347468; Foxb2.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS50039; FORK_HEAD_3; 1.
KW DNA-binding; Nuclear protein; Transcription regulation.
FT DNA_BIND 12 103 FORK-HEAD.
FT DOMAIN 139 153 POLY-HIS.
FT DOMAIN 156 162 POLY-HIS.
FT DOMAIN 163 172 POLY-PRO.
FT DOMAIN 217 231 POLY-PRO.
FT DOMAIN 249 258 POLY-ALA.
FT DOMAIN 321 330 POLY-ALA.
FT DOMAIN 396 399 POLY-ALA.
SQ SEQUENCE 428 AA; 45170 MW; DBA8EFD1E94AB10 CRC64;

Query Match 15.9%; Score 154; DB 1; Length 428;
Best Local Similarity 29.5%; Pred. No. 1.2e-05;
Matches 52; Conservative 6; Mismatches 66; Indels 52; Gaps 10;

QY 1 HPHKHSHEQHPHGHHPAH-HPHEHDTHQHPHGHHPHGHHPHGHHPH-GHHPHGHHPH 58
DQ 116 HAHLSGSSKGAPGTGPGHLPH-----HPH---HAHHHHHHHHAAHHHHHP- 163

QY 59 CHDFQYGPDCPPH--NQHCCHGHGPPPHGLRRRGKGRP----- 100
DQ 164 -----PQPPPPPPHVPYFHQPAPAPQPPHPSQ-PAQQPQSQPQTSHPGKMQEA 217

QY 101 -----FHCROIGSVYRL--PPLRKGEVLPPEANFPSPFPLPHKHPLKPDN 144
DQ 218 AAVAAAAAAGAAVGVSLQSPYGLGSAAGAAAAAAS--TTGFKHPFAIEN 271

RESULT 13
HMN1_DROME STANDARD; PRT; 659 AA.
AC P22807; Q9VD96;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Homeobox protein slou (S59/2) (slouch protein) (Homeobox protein NK-1).
DE 1).
GN SLOU OR NK1 OR S59 OR CG6534.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91099659; PubMed=1980118;
RA Dohrmann C., Azpiazu N., Frasch M.;
RT "A new Drosophila homeo box gene is expressed in mesodermal precursor cells of distinct muscles during embryogenesis.";
RL Genes Dev. 4:2098-2111(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [3]
RP REVISIONS.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a systematic review.";
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN [4]
RP SEQUENCE OF 497-625 FROM N.A.
RX MEDLINE=90046666; PubMed=2573058;
RA Kim Y., Nirenberg M.;
RT "Drosophila NK-homeobox genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:7716-7720(1989).
CC -!- FUNCTION: May play a role in specifying the identity of particular somatic muscles and neurons of the CNS.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- TISSUE SPECIFICITY: Mesodermal precursor cells of distinct muscles during embryogenesis, a subset of neuronal cells of the CNS and their precursors and also in cells of a small region of the midgut.
CC -!- DEVELOPMENTAL STAGE: Postgastrulation-stage.
CC -!- SIMILARITY: Belongs to the NK-1 homeobox family.
CC -!- SIMILARITY: Contains 1 homeobox domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X55393; CAA39067.1; --
CC EMBL; AE003735; AAF5901.3; --
CC EMBL; M27289; AAA28616.1; --
CC PIR; A36664; A36664.
CC HSSP; P14653; 1B72.
CC TRANSFAC; T04257; --

DR FlyBase; FBgn0002941; slou.
DR GO: 0005634; C:nucleus; NAS.
DR GO: 0003702; F:RNA polymerase II transcription factor acti. . .; NAS.
DR GO: 0007501; P:mesoderm cell fate specification; NAS.
DR GO: 0007521; P:muscle cell fate determination; IEP.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom; PD00010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX 1; 1.
DR PROSITE; PS50071; HOMEBOX 2; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
Repeat.
FT DOMAIN 201 239 HIS-RICH.
FT DOMAIN 221 234 7 X 2 AA TANDEM REPEATS OF H-P.
FT DOMAIN 364 372 POLY-ALA.
FT DOMAIN 477 522 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 536 542 POLY-GLY.
FT DNA_BIND 545 604 HOMEBOX.
SQ SEQUENCE 659 AA; 69955 MW; 5D401F55C4670280 CRC64;

Query Match 15.5%; Score 149.5; DB 1; Length 659;
Best Local Similarity 30.6%; Pred. No. 4.2e-05;
Matches 44; Conservative 4; Mismatches 49; Indels 47; Gaps 9;

QY 10 QHPHG---HHPH-AHHPHEHDTHQHPHGHHHPGHHPGH-HPHGH-HPHCHDF 62
Db 175 QHPHALLQHPHLLQNPFLAAQHHM-HHQHQHHPHHPHGHHPHHPH----- 229
QY 63 QDYGCDPPPHNQCHCHGCHGPPPPGHLRRRGKGPGRPFHCRQIGSVYRLPLRKGEVLP 122
Db 230 -----PHPHPSA-----VFHLRAPSSTAPP--SPATSP 257
QY 123 LPEANFPSPFLPHHKHPLKPDNPQ 146
Db 258 LSP---PTSPAMHSDQMSPPPIAP 278

RESULT 14
PYGO DROME STANDARD; PRT; 815 AA.
AC Q9V9W8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Pygopus protein (Gammy legs protein).
GN PYGO OR GAM OR CG11518.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21952490; PubMed=11955446;
RA Kramps T., Peter O., Brunner E., Nellen D., Froesch B., Chatterjee S.,
RA Murone M., Zuellig S., Basler K.;
RT "Wnt/wingless signaling requires BCL9/legless-mediated recruitment of
pygopus to the nuclear beta-catenin-TCF complex."
RL Cell 109:47-60 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Head;
RX MEDLINE=22010046; PubMed=12015286;
RA Parker D.S., Jemison J., Cadigan K.M.;
RT "Pygopus, a nuclear PHD-finger protein required for wingless signaling
in Drosophila."
RL Development 129:2565-2576 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley; TISSUE=Embryo;
RX MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA Rubin G.M., Celniker S.E.;
RT "A Drosophila full-length cDNA resource."
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8 (2002).
CC -!- FUNCTION: Involved in signal transduction through the Wnt pathway.
CC -!- SUBUNIT: Binds to BCL9 via the PHD-type zinc finger motif, and
CC thereby becomes part of the nuclear ARM/PAN complex.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: Ubiquitous throughout embryogenesis and larval
CC development.
CC -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically
CC throughout development.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF457206; AAL91369.1; -
CC EMBL; AY075095; AAL79357.1; -
CC EMBL; AE003778; AAF57161.1; -
CC EMBL; AY058500; AAL13729.1; -
CC FlyBase; FBgn0043900; pygo.
CC GO: 0005634; C:nucleus; NAS.
CC GO: 00030528; F:transcription regulator activity; IPI.
CC GO: 00030177; P:positive regulation of Wnt receptor signal. . .; IPI.

FT	ACT SITE	978	978	BY SIMILARITY.
FT	BINDING	924	924	SUBSTRATE (BY SIMILARITY).
FT	BINDING	979	979	SUBSTRATE (BY SIMILARITY).
FT	DISULFID	627	880	BY SIMILARITY.
FT	DISULFID	776	789	BY SIMILARITY.
FT	DISULFID	799	822	BY SIMILARITY.
FT	DISULFID	806	815	BY SIMILARITY.
FT	DISULFID	844	851	BY SIMILARITY.
FT	CARBOHYD	659	659	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	MUTAGEN	627	627	C->T: 36% OF ORIGINAL ACTIVITY.
SQ	SEQUENCE	1002 AA;	114237 MW;	4A8D81CFDAB2D854 CRC64;

Query Match 15.5%; Score 149.5; DB 1; Length 1002;
Best Local Similarity 31.4%; Pred. No. 6.1e-05;
Matches 49; Conservative 5; Mismatches 63; Indels 39; Gaps 10;

QY	2	PHKHSHEQH----	PHGHPHAPH-----	HEHDTHQPHGHHPGHGHH----	PHGHPHG	49
Db	210	PPMHHKPGEHMPPPP	MHEPGEHMPPPP	MHEPGEHMPPPP	MHEPGEHMPPPP	269
QY	50	HH-----	PHGHPHCHDFQDYGPCDPPP	--	HNQCHCCHGHGPPPGHLLRRRGPGKGRPFHC	103
Db	270	EHMPPPPMHHEPGEH	-----	MPPPPMHHEPGE-----	HMPPPPMHHEPGEHMPPPPMH-	317
QY	104	RQIGSVYRLPPL--	RKGEVLPLPEANFP	SFPLPHHK	137	
Db	318	HEPGEHMPPPPMHHEPGEHMPPPP	-----	PFKHHE	346	

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2004, 22:25:45 ; Search time 87.3735 Seconds
(without alignments)
534.449 Million cell updates/sec

Title: US-10-074-225A-5
Perfect score: 966
Sequence: 1 HPHKHSHSEQHPGHHPAH.....PSFPLPHKHPLKPDNQPEP 148

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	645.5	66.8	525	11 Q99PS6	Q99ps6 mus musculu
2	645.5	66.8	525	11 Q9ESB3	Q9esb3 mus musculu
3	645.5	66.8	525	11 Q99PS5	Q99ps5 mus musculu
4	644	66.7	525	11 Q99PS8	Q99ps8 rattus norv
5	644	66.7	546	6 Q9BGU1	Q9bgul bos taurus
6	626	64.8	515	11 Q99PS7	Q99ps7 rattus norv
7	614.5	63.6	510	11 Q9ESB2	Q9esb2 rattus norv
8	281.5	29.1	233	5 Q8MP30	Q8mp30 dictyosteli
9	223.5	23.1	294	2 Q9KI87	Q9ki87 bacillus ce
10	217	22.5	140	5 Q26056	Q26056 plasmodium
11	216.5	22.4	296	2 Q9KI89	Q9ki89 bacillus ce
12	213.5	22.1	465	13 Q801E5	Q80le5 xenopus lae
13	205.5	21.3	273	2 Q9KI88	Q9ki88 bacillus ce
14	205.5	21.3	462	13 Q7ZY91	Q7zy91 xenopus lae
15	205.5	21.3	462	13 Q7SYH2	Q7syh2 xenopus lae
16	205	21.2	295	16 Q816T1	Q816t1 bacillus ce

17	204.5	21.2	479	5 Q8TOB3	Q8t0b3 drosophila
18	204.5	21.2	783	5 Q9VSZ7	Q9vsz7 drosophila
19	199.5	20.7	130	2 Q9KI77	Q9ki77 bacillus an
20	199.5	20.7	133	2 Q9KI76	Q9ki76 bacillus an
21	199.5	20.7	265	2 Q9KI91	Q9ki91 bacillus an
22	195	20.2	112	11 Q9CR74	Q9cr74 mus musculu
23	193	20.0	735	5 Q9NES7	Q9nes7 caenorhabdi
24	191.5	19.8	341	5 Q9VV44	Q9vv44 drosophila
25	191.5	19.8	485	5 Q76876	Q76876 drosophila
26	188	19.5	403	16 Q98H88	Q98h88 rhizobium 1
27	182	18.8	118	2 Q9KI82	Q9ki82 bacillus an
28	182	18.8	121	2 Q9KI80	Q9ki80 bacillus an
29	182	18.8	124	2 Q9KI78	Q9ki78 bacillus an
30	182	18.8	124	2 Q9KI79	Q9ki79 bacillus an
31	182	18.8	247	16 Q8IKK8	Q8ikk8 bacillus an
32	179.5	18.6	599	13 Q7SXG7	Q7sxcg7 brachydanio
33	179.5	18.6	1264	5 Q86H34	Q86h34 dictyosteli
34	179	18.5	372	16 Q89CP3	Q89cp3 bradyrhizob
35	179	18.5	554	5 Q9W4C1	Q9w4c1 drosophila
36	178.5	18.5	118	2 Q9KI81	Q9ki81 bacillus an
37	177	18.3	115	2 Q9KI83	Q9ki83 bacillus an
38	177	18.3	356	13 Q98UK5	Q98uk5 brachydanio
39	177	18.3	356	13 Q73679	Q73679 brachydanio
40	177	18.3	1572	11 Q8R5N0	Q8r5n0 mus musculu
41	177	18.3	1617	11 Q80U47	Q80u47 mus musculu
42	176.5	18.3	443	5 Q8IH14	Q8ih14 drosophila
43	176.5	18.3	2148	5 Q8IR53	Q8ir53 drosophila
44	176.5	18.3	2196	5 Q9VY40	Q9vy40 drosophila
45	176	18.2	623	5 P90533	P90533 dictyosteli

ALIGNMENTS

RESULT 1

Q99PS6 PRELIMINARY; PRT; 525 AA.
AC Q99PS6;
DT 01-JUN-2001 (TREMELrel. 17, Created)
DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Histidine-rich glycoprotein.
GN AW413091 OR MTHRG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RA Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,
RA Ohashi T., Sato N., Hirata D., Tsuchida N., Koide T.;
RT "Molecular diversity of mammalian histidine-rich glycoprotein.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB055897; BAB33094.1; -.
DR MGD; MGI:2146636; AW413091.
DR GO; GO:0004869; F:cysteine protease inhibitor activity; IEA.
DR InterPro; IPR000010; Cystatin.
DR Pfam; PF00031; cystatin; 1.
DR SMART; SM00043; CY; 2.
SQ SEQUENCE 525 AA; 59090 MW; A83E93A439CFB3AC CRC64;

Query Match 66.8%; Score 645.5; DB 11; Length 525;
Best Local Similarity 69.5%; Pred. No. 7.9e-56;
Matches 105; Conservative 9; Mismatches 34; Indels 3; Gaps 1;

Qy	1	HPKHHSHEQHPGHHPAHHPHEDTHRQHPGHHPGHHPGHHPGHHPGHHPGH	60
Db	347	HPSHHPGHHSHGHHPGHHPGHSHSHGHHPGHHPGHHPGHHPGHHPGH	406
Qy	61	DFQDYGPCDPPPHNQ---GHCHGHGPPPPGHLRRRGPGKGRPFHCRQIGSVYVLPPLRK	117
Db	407	DFLDYGPCDPPSPNSQELKGQYHRYGPPGHGHSRKRGPCKGLFPFHHQIGYVYVLPPLNI	466

QY 118 GEVLPPLPEANFPSPFLPHHKLKPDNQPPF 148
||||| ||||| ||||| ||||| :||: |||||
Db 467 GEVLTLPPEANFPSPFLPNCNRSRSLQPEIQPPF 497

RESULT 2

Q9ESB3
ID Q9ESB3 PRELIMINARY; PRT; 525 AA.

DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Histidine-rich glycoprotein.

GN AW413091 OR HRG.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=129;

RX MEDLINE=20307726; PubMed=10849117;

RA Hulett M.D., Parish C.R.;

RT "Murine histidine-rich glycoprotein: cloning, characterization and

cellular origin";

RL Immunol. Cell Biol. 78:280-287(2000).

DR EMBL; AF194028; AAG28416.1; -.

DR MGD; MGI:2146636; AW413091.

DR GO; GO:0004869; F:cysteine protease inhibitor activity; IEA.

DR InterPro; IPR000010; Cystatin.

DR Pfam; PF00031; cystatin; 1.

DR SMART; SM00043; CY; 2.

SQ SEQUENCE 525 AA; 59132 MW; 6E55F2A439CFB123 CRC64;

Query Match 66.8%; Score 645.5; DB 11; Length 525;

Best Local Similarity 69.5%; Pred. No. 7.9e-56;

Matches 105; Conservative 9; Mismatches 34; Indels 3; Gaps 1;

QY 1 HPHKHSHEQHPHGHHPHHAHPHEHDTHQHPHGHHPHGHHPHGHHPHGHHPHGH 60

||||| ||||| ||||| ||||| :||: ||||| ||||| ||||| ||||| ||||| |||||

Db 347 HPHSHHPGHHSHGHHPHGHHPHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSH 406

QY 61 DFQYGPCCDPPPHNQ---GHCCHGHGPPPGHLLRRRGKGRPFHCRQIGSVYRLPLRK 117

||||| ||||| ||||| ||||| :||: ||||| ||||| ||||| ||||| ||||| |||||

Db 407 DFLDYGPCDPPSNSQELKGQYHRGYGPPHGHSHSHSHSHSHSHSHSHSHSHSH 466

QY 118 GEVLPPLPEANFPSPFLPHHKLKPDNQPPF 148

||||| ||||| ||||| ||||| :||: |||||

Db 467 GEVLTLPPEANFPSPFLPNCNRSRSLQPEIQPPF 497

RESULT 3

Q99PS5

ID Q99PS5 PRELIMINARY; PRT; 525 AA.

AC Q99PS5;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Histidine-rich glycoprotein.

GN AW413091 OR MHRG.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA Tsuchida N., Wakabayashi S., Jahn-Dechent W., Koide T.;

RT "Structure of mouse histidine-rich glycoprotein gene";

RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Liver;

RA Strausberg R.;

RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB055898; BAB33095.1; -.

DR EMBL; BC011168; AAH11168.1; -.

DR MGD; MGI:2146636; AW413091.

DR GO; GO:0004869; F:cysteine protease inhibitor activity; IEA.

DR InterPro; IPR000010; Cystatin.

DR Pfam; PF00031; cystatin; 1.

DR SMART; SM00043; CY; 2.

SQ SEQUENCE 525 AA; 59162 MW; A83E93A439CFB126 CRC64;

Query Match

Best Local Similarity 66.8%; Score 645.5; DB 11; Length 525;

Matches 105; Conservative 9; Mismatches 34; Indels 3; Gaps 1;

QY 1 HPHKHSHEQHPHGHHPHHAHPHEHDTHQHPHGHHPHGHHPHGHHPHGHHPHGH 60

||||| ||||| ||||| ||||| :||: ||||| ||||| ||||| ||||| ||||| |||||

Db 347 HPHSHHPGHHSHGHHPHGHHPHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSH 406

QY 61 DFQYGPCCDPPPHNQ---GHCCHGHGPPPGHLLRRRGKGRPFHCRQIGSVYRLPLRK 117

||||| ||||| ||||| ||||| :||: ||||| ||||| ||||| ||||| ||||| |||||

Db 407 DFLDYGPCDPPSNSQELKGQYHRGYGPPHGHSHSHSHSHSHSHSHSHSHSHSH 466

QY 118 GEVLPPLPEANFPSPFLPHHKLKPDNQPPF 148

||||| ||||| ||||| ||||| :||: |||||

Db 467 GEVLTLPPEANFPSPFLPNCNRSRSLQPEIQPPF 497

RESULT 4

Q99PS8

ID Q99PS8 PRELIMINARY; PRT; 525 AA.

AC Q99PS8;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Histidine-rich glycoprotein 1.

GN RNRHG1.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Sprague-Dawley; TISSUE=Liver;

RA Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,

Ohashi T., Sato N., Hirata D., Tsuchida N., Koide T.;

"Molecular diversity of mammalian histidine-rich glycoprotein.";

RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB055895; BAB33092.1; -.

DR GO; GO:0004869; F:cysteine protease inhibitor activity; IEA.

DR InterPro; IPR000010; Cystatin.

DR Pfam; PF00031; cystatin; 1.

DR SMART; SM00043; CY; 2.

SQ SEQUENCE 525 AA; 59049 MW; 38290A631FAC7777 CRC64;

Query Match

Best Local Similarity 66.7%; Score 644; DB 11; Length 525;

Matches 107; Conservative 7; Mismatches 34; Indels 8; Gaps 2;

QY 1 HPHKHSHEQHPHGHHPHHAHPHEHDTHQHPHGHHPHGHHPHGHHPHGHHPHGH 55

||||| ||||| ||||| ||||| :||: ||||| ||||| ||||| ||||| ||||| |||||

Db 342 HPHGHHPGQHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPGH 401

QY 56 HPHCHDFQDYGPCDPPPHNQ---GHCCHGHGPPPGHLLRRRGKGRPFHCRQIGSVYRL 112

||||| ||||| ||||| ||||| :||: ||||| ||||| ||||| ||||| ||||| |||||

Db 402 HPHGHDFLDYGPCDPPSNSQELKGQYHRGYGPPHGHSHSHSHSHSHSHSHSHSH 461

QY 113 PPLRKGEVLPPEANFPSPFLPHHKLKPDNQPPF 148

||||| ||||| ||||| ||||| :||: |||||

Db 462 PPLNVGEVLTPEANFPSPFLPNCNRSRSLQPEIQPPF 497

RESULT 5

Q9BGU1 PRELIMINARY; PRT; 546 AA.
ID Q9BGU1;
AC Q9BGU1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Histidine-rich glycoprotein.
GN BTHRG.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,
RA Ohashi T., Sato N., Hirata D., Tsuchida N., Koide T.;
RT "Molecular diversity of mammalian histidine-rich glycoprotein.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB055894; BAB33091.1; -.
DR GO; GO:0004869; F:cysteine protease inhibitor activity; IEA.
DR InterPro; IPR00010; Cystatin.
DR Pfam; PF00031; cystatin; 1.
DR SMART; SM00043; CY; 2.
SQ SEQUENCE 546 AA; 61948 MW; 26264858824D89EE CRC64;

Query Match 66.7%; Score 644; DB 6; Length 546;
Best Local Similarity 62.7%; Pred. No. 1.1e-55;
Matches 104; Conservative 10; Mismatches 34; Indels 18; Gaps 2;

QY 1 HPHKHSHEQHPGHHPHAHHPHEHDTTHRQHPGHHPGHHPGHHPGHHPGHHPCH 60
Db 353 HPHGHHPGHHPGHHPGHHPGHHPGHHPGHHPGHHPGHHPGHHPGHHPGH 412
QY 61 -----DFQDYGPCDPPPHN---QGHCHGHGPPPGHLLRRRGPGKGRPFH 102
Db 413 HPYGHHPGHHPHDNDFYDHGCDPPPHRQDPQDHRQGRDPPPRHSKERGPGKGRFV 472
QY 103 CRQIGSVYRLPPLRKGEVLPPEANFSPFPLPHKHPKLPKPDNQPPF 148
Db 473 WRPTGYIHLPSLKKGEVLPPEANFSPFSLPNHNNPLQPEIQAFP 518

RESULT 6
Q99PS7 PRELIMINARY; PRT; 515 AA.
ID Q99PS7;
AC Q99PS7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Histidine-rich glycoprotein 2.
GN RNHRG2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RA Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,
RA Ohashi T., Sato N., Hirata D., Tsuchida N., Koide T.;
RT "Molecular diversity of mammalian histidine-rich glycoprotein.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB055896; BAB33093.1; -.
DR GO; GO:0004869; F:cysteine protease inhibitor activity; IEA.
DR InterPro; IPR00010; Cystatin.
DR Pfam; PF00031; cystatin; 1.
DR SMART; SM00043; CY; 2.
SQ SEQUENCE 515 AA; 58055 MW; 7CEBA3A1A3678966 CRC64;

Query Match 64.8%; Score 626; DB 11; Length 515;
Best Local Similarity 69.5%; Pred. No. 6.5e-54;
Matches 105; Conservative 6; Mismatches 32; Indels 8; Gaps 2;

QY 1 HPHKHSHEQHPGHHPHAHHPHEHDTTHRQHPGHHPGHHPGHHPGHHPGHHPCH 60
Db 342 HPHGHHPGHHPGHHPGHHPGHHPGHHPGHHPGHHPGHHPGHHPGHHPGH 396
QY 61 DFQDYGPCDPPPHNQ---GHCHGHGPPPGHLLRRRGPGKGRPFHCRQIGSVYRLPPLRK 117
Db 397 DFLDYGPCDPPPSNSQELKGVHRGHGPPPHGHSRKRGPCKGLFFPHQRQIGYVYRLPPLNV 456
QY 118 GEVLPPLPEANFSPFPLPHKHPKLPKPDNQPPF 148
Db 457 GEVLTPEANFSPFSLPNCNRPQPPEIQPPF 487

RESULT 7
Q9ESB2 PRELIMINARY; PRT; 510 AA.
ID Q9ESB2
AC Q9ESB2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Histidine-rich glycoprotein.
GN HRG.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Lewis;
RX MEDLINE=20307726; PubMed=10849117;
RA Hulett M.D., Parish C.R.;
RT "Murine histidine-rich glycoprotein: cloning, characterization and
RT cellular origin."
RL Immunol. Cell Biol. 78:280-287(2000).
DR EMBL; AF194029; AAG28417.1; -.
DR GO; GO:0004869; F:cysteine protease inhibitor activity; IEA.
DR InterPro; IPR00010; Cystatin.
DR Pfam; PF00031; cystatin; 1.
DR SMART; SM00043; CY; 2.
SQ SEQUENCE 510 AA; 57581 MW; 508E6E06AA2ED58E CRC64;

Query Match 63.6%; Score 614.5; DB 11; Length 510;
Best Local Similarity 68.9%; Pred. No. 8.9e-53;
Matches 102; Conservative 7; Mismatches 32; Indels 7; Gaps 2;

QY 1 HPHKHSHEQHPGHHPHAHHPHEHDTTHRQHPGHHPGHHPGHHPGHHPGHHPCH 60
Db 342 HPHGHHPGHHPGHHPGHHPGHHPGHHPGHHPGHHPGHHPGHHPGHHPGH 396
QY 61 DFQDYGPCDPPPHNQCHCHGHGPPPGHLLRRRGPGKGRPFHCRQIGSVYRLPPLRKGEV 120
Db 397 DFLDYGPCDPPPSNSQYH--QGHGPPHGSRRKRGPCKGLFFPHQRQIGYVYRLPPLNVGEV 454
QY 121 LPLPEANFSPFPLPHKHPKLPKPDNQPPF 148
Db 455 LTPPEANFPIFSLPNCNRPQPPEIRPPF 482

RESULT 8
Q8MP30 PRELIMINARY; PRT; 233 AA.
ID Q8MP30
AC Q8MP30; Q8T164;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein (Similar to plasmodium lophurae histidine-rich
DE glycoprotein).
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=AX4;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=AX4;
RX MEDLINE=22092622; PubMed=12097910;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and analysis of chromosome 2 of Dictyostelium discoideum.";
RL Nature 418:79-85(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Baumgart C.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC123513; AAM44363.1; -;
DR EMBL; AC117070; AAM09303.2; -;
KW Hypothetical protein.
SQ SEQUENCE 233 AA; 28937 MW; AD4C7A2F86BB14B9 CRC64;

Query Match 29.1%; Score 281.5; DB 5; Length 233;
Best Local Similarity 38.3%; Pred. No. 3.6e-20;
Matches 57; Conservative 3; Mismatches 62; Indels 27; Gaps 5;

QY 1 HPHKHSHEQHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHCH 60
Db 79 HHH 138

QY 61 DFQDYGCDPPPHNQHCCHGHPGPPPGHLRRRGPKGPRPFHCRQIGSVYRLPPLRKGEV 120
Db 139 HHH-----HH 173

QY 121 LPLPEANFPSPFLPH-HKHPLKPDNQPP 148
Db 174 NPHEPH 201

RESULT 9
QY 94
ID Q9KI87 PRELIMINARY; PRT; 294 AA.
AC Q9KI87;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative VrrB (Fragment).
GN VRRB.
OS Bacillus cereus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1396;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 43881;
RX MEDLINE=20327579; PubMed=10869077;
RA Schupp J.M., Klevytska A.M., Zinser G., Price L.B., Keim P.;
RT "vrrB, a hypervariable open reading frame in bacillus anthracis.";
RL J. Bacteriol. 182:3989-3997(2000).
DR EMBL; AF238888; AAF86203.1; -;
FT NON TER 1
SQ SEQUENCE 294 AA; 30753 MW; C05B7B37FE2E395C CRC64;

Query Match 23.1%; Score 223.5; DB 2; Length 294;
Best Local Similarity 38.3%; Pred. No. 2.4e-14;
Matches 49; Conservative 5; Mismatches 41; Indels 33; Gaps 6;

QY 1 HPHKHSHEQHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHCH 49
Db 118 HGHGHH 177

QY 50 HH-----PHGHHPHCHDFQDYGCDPPPHNQHCCH-CHGPPPPGHLRRGP 94
Db 178 HHHHPQAVLYQTHQGQGHGHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH 230
QY 95 GKGPFPFH 102
Db 231 HHGHQGH 238

RESULT 10
QY 94
ID Q26056 PRELIMINARY; PRT; 140 AA.
AC Q26056;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Histidine-rich protein (Fragment).
OS Plasmodium lophurae.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5853;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86174893; PubMed=3007981;
RA Irving D.O., Cross G.A.M., Feder R., Wallach M.;
RT "Structure and organization of the histidine-rich protein gene of
RL Mol. Biochem. Parasitol. 18:223-234(1986).
DR EMBL; M15317; AAA29616.1; -;
DR PIR; A54523; A54523.
FT NON TER 1
SQ SEQUENCE 140 AA; 18216 MW; ED05EED8C805B2FA CRC64;

Query Match 22.5%; Score 217; DB 5; Length 140;
Best Local Similarity 39.6%; Pred. No. 5.3e-14;
Matches 38; Conservative 2; Mismatches 34; Indels 22; Gaps 2;

QY 2 PHKHSHEQHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHCH 46
Db 8 PHHH 67

QY 47 PHGHHPHGHHPHCHDFQDYGCDPPPHNQHCCHG 82
Db 68 HHHHHHPWFHH 96

RESULT 11
QY 94
ID Q9KI89 PRELIMINARY; PRT; 296 AA.
AC Q9KI89;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative VrrB (Fragment).
GN VRRB.
OS Bacillus cereus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1396;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 11778;
RX MEDLINE=20327579; PubMed=10869077;
RA Schupp J.M., Klevytska A.M., Zinser G., Price L.B., Keim P.;
RT "vrrB, a hypervariable open reading frame in bacillus anthracis.";
RL J. Bacteriol. 182:3989-3997(2000).
DR EMBL; AF238888; AAF86201.1; -;
FT NON TER 1
SQ SEQUENCE 296 AA; 31360 MW; 332D6BC6E0A57A5A CRC64;

Query Match 22.4%; Score 216.5; DB 2; Length 296;
Best Local Similarity 37.3%; Pred. No. 1.2e-13;
Matches 50; Conservative 4; Mismatches 35; Indels 45; Gaps 8;

QY 1 HPHKHSHEQHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHCH 47

```

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 31293;
RX MEDLINE=20327579; PubMed=10869077;
RA Schupp J.M., Klevytska A.M., Zinser G., Price L.B., Keim P.;
RT "vrrB, a hypervariable open reading frame in bacillus anthracis.";
RL J. Bacteriol. 182:3989-3997(2000).
FR EMBL; AF238887; AAF86202.1; -.
FT NON_TER
SQ SEQUENCE 273 AA; 28573 MW; B80F2979612BEB0C CRC64;

Query Match      21.3%; Score 205.5; DB 2; Length 273;
Best Local Similarity 36.9%; Pred. No. 1.3e-12;
Matches 45; Conservative 3; Mismatches 27; Indels 47; Gaps 7;

QY 3 HKHSHSQHPHGHHPHAHHPHEHDTHQPH-GHH-----PHGHH-----PHGH----- 45
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 119 HGHHHHQCGQHGHGHHHHHQQGQHGHGHHHHHQQGHHHQQGSHHGHHHHHQQGHGHHGQQV 178
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 46 HPHGHH-----PHGHHPHCHDFQDYGPCDPPPHNQGHCHCHGPPPGHLRR 91
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 179 HHGHHHHHPQAVLYQTHQGHGHHGH-----HHQGH--HHHGHGHHGHHG 222
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 92 RG 93
   :|
Db 223 QG 224

RESULT 14
Q7ZY91 PRELIMINARY; PRT; 462 AA.
AC Q7ZY91;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to fetuin B.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBDJ databases.
DR EMBL; BC043891; AAH43891.1; -.
DR GO; GO:0004869; F:cysteine protease inhibitor activity; IEA.
DR InterPro; IPR000010; Cystatin.
DR Pfam; PF00031; cystatin; 2.
DR SMART; SM00043; CY; 2.
SQ SEQUENCE 462 AA; 53185 MW; D7BAD339961739FB CRC64;

Query Match      21.3%; Score 205.5; DB 13; Length 462;
Best Local Similarity 32.5%; Pred. No. 2.2e-12;
Matches 49; Conservative 12; Mismatches 65; Indels 25; Gaps 6;

QY 1 HPHKHH-SHEQHPGHHHPAHHPHEHDTHQPHGHHHPHGHHHPHGHHHPHGHHHPHC 59
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 312 HKRHHHPGSHKGKHGRHHHHHPHHHD-----HPPHHHHHHHHHPNHTSSSEHHEHH 363
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 60 HDFQDYGPCDPPPHNQGHCHCHG-PPGHLRRGPGKGRP-FHCRQIGSVYVLPPLRK 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 364 H-----HHPNHTSSSEHGSSEHTDKAGKKADRCFMTSRKGVVQKITLASE 411
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 118 GEVLPPLPEANF--PSFPLPHKKHPLKPDNQ 146
   :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 412 TDVLPAPTITISRPSRTEYIQFPEAASHLP 442

RESULT 15
Q7SYH2 PRELIMINARY; PRT; 462 AA.
ID Q7SYH2
AC Q7SYH2;

```


FILE REFERENCE: Hutchens
CURRENT APPLICATION NUMBER: US/08/785,636
CURRENT FILING DATE: 1997-01-17
EARLIER APPLICATION NUMBER: 08/068,896
EARLIER FILING DATE: 1993-05-28
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 26
TYPE: PRT
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: Description of Unknown Organism: peptide
US-08-785-636-1

Query Match 19.2%; Score 185; DB 3; Length 26;
Best Local Similarity 100.0%; Pred. No. 6.9e-12;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 34 GHHPHGHHPHGHHPHGHHPHGHHPH 58
Db 1 GHHPHGHHPHGHHPHGHHPHGHHPH 25

RESULT 3
US-09-095-407-2
Sequence 2, Application US/09095407
Patent No. 6124137
GENERAL INFORMATION:
APPLICANT: Hutchens, T. William
TITLE OF INVENTION: Surface-Enhanced Neat Desorption for the Detection of Analytes

NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fulbright & Jaworski L.L.P.
STREET: 1301 McKinney, Suite 5100
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77010-3095
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/095,407
FILING DATE: 10-Jun-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/068,896
FILING DATE: 28-MAY-1993
APPLICATION NUMBER: WO PCT/US94/06064
FILING DATE: 27-MAY-1994
APPLICATION NUMBER: US 08/556,951
FILING DATE: 27-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Paul, Thomas D.
REGISTRATION NUMBER: 32,714
REFERENCE/DOCKET NUMBER: D-5639-PCT-US-D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (713) 651-5325
TELEFAX: (713) 651-5246
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 26 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-095-407-2

Query Match 19.2%; Score 185; DB 3; Length 26;
Best Local Similarity 100.0%; Pred. No. 6.9e-12;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 34 GHHPHGHHPHGHHPHGHHPHGHHPH 58
Db 1 GHHPHGHHPHGHHPHGHHPHGHHPH 25

RESULT 4
US-09-742-494-2
Sequence 2, Application US/09742494
Patent No. 6528320
GENERAL INFORMATION:
APPLICANT: HUTCHENS, WILLIAM
APPLICANT: YIP, TAI-TUNG
TITLE OF INVENTION: METHODS AND APPARATUS FOR DESORPTION AND IONIZATION OF ANALYTES
FILE REFERENCE: P00798USE/09306611
CURRENT APPLICATION NUMBER: US/09/742,494
CURRENT FILING DATE: 2000-12-20
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2
LENGTH: 26
TYPE: PRT
ORGANISM: Homo sapiens
US-09-742-494-2

Query Match 19.2%; Score 185; DB 4; Length 26;
Best Local Similarity 100.0%; Pred. No. 6.9e-12;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 34 GHHPHGHHPHGHHPHGHHPHGHHPH 58
Db 1 GHHPHGHHPHGHHPHGHHPHGHHPH 25

RESULT 5
US-09-507-323B-13
Sequence 13, Application US/09507323B
Patent No. 6274345
GENERAL INFORMATION:
APPLICANT: Choi, Jong Hyun
APPLICANT: Lee, Sang Yup
APPLICANT: Xu, Zhaochui
TITLE OF INVENTION: EXPRESSION VECTORS COMPRISING A GENE
TITLE OF INVENTION: ENCODING OMPC DERIVED FROM E.COLI AS CELL SURFACE ANCHORING
TITLE OF INVENTION: MOTIF
FILE REFERENCE: HYLEE39.001AUS
CURRENT APPLICATION NUMBER: US/09/507,323B
CURRENT FILING DATE: 2000-02-18
PRIOR APPLICATION NUMBER: KR 10-1999-0005773
PRIOR FILING DATE: 1999-02-22
NUMBER OF SEQ ID NOS: 13
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13
LENGTH: 167
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Histidine linker.
US-09-507-323B-13

Query Match 19.1%; Score 184.5; DB 3; Length 167;
Best Local Similarity 32.7%; Pred. No. 5.6e-11;
Matches 48; Conservative 2; Mismatches 50; Indels 47; Gaps 9;
QY 1 HPHKHHS-----HEQH-----PHGHH-----PHAHHPHED-----THROH--- 31
Db 8 HHHHHHSGLDPSGHHHHHSGLDPSGHHHHHSGLDPSGHHHHHSGLDPSGHHHHHSG 67
QY 32 --PHGHHPHGH---HPHGHHPHGH---HPHGHHPHCHDFOD-----YGPCDPPPHNQ 75

TOZ-60-60

	Query Match	16.1%;	Score 155.5;	DB 4;	Length 865;
	Best Local Similarity	29.2%;	Pred. No. 2.3e-07;		
Matches	47; Conservative	9;	Mismatches 88;	Indels 17;	Gaps 6;
QY	2	PKKHSHEQHGHHPHAAHHPEHEDHTRQHPHGHHPHGHHHPHGHHHPGHCHD	61		
Dδ	571	PDRQHPPHPPRRNPPEASPPESPNNWQHPPHPPRPNNPPEASPPESPNNWQHPPHP	630		
QY	62	FQDYGP--CDPP---PHNQGHCCGHGGPPPGHLRRRCCKGP----	RPF-----HCRQIGSV	109	
Dδ	631	PRNPPGASPPESPNNWQH-PHPHPPRRPNPPEASPPQSSPPPEPQRPFQSPHTEFFHY	689		
QY	110	YRLP----PLRKGEVLPLPEANFPSPFPLPHHKHPLKPDNQP	146		
Dδ	690	HPYPGNLPFYFTYHQSPFLPYGYGRDPCPCASHYPADDS	P	730	

```

RESULT 10
US-09-507-323B-12
; Sequence 12, Application US/09507323B
; Patent No. 6274345
; GENERAL INFORMATION:
; APPLICANT: Choi, Jong Hyun
; APPLICANT: Lee, Sang Yup
; APPLICANT: Xu, Zhaohui
; TITLE OF INVENTION: EXPRESSION VECTORS COMPRISING A GENE
; TITLE OF INVENTION: ENCODING OMPC DERIVED FROM E.COLI AS
; TITLE OF INVENTION: MOTIF
; FILE REFERENCE: HYLEE39.001AUS
; CURRENT APPLICATION NUMBER: US/09/507,323B
; CURRENT FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: KR 10-1999-0005773
; PRIOR FILING DATE: 1999-02-22
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Histidine linker.
US-09-507-323B-12

```

```

Query Match      15.4%; Score 148.5; DB 3; Length 89;
Best Local Similarity 42.5%; Pred. No. 9.9e-08;
Matches 31; Conservative 0; Mismatches 27; Indels 15; Gaps 4;

QY 1 HPHKHHS-----HEQHPHGCHPHAAHPHEHDTTRQHPHGHHPHGH---HPHGHHPHGH 50
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8 HHHHHHSGLDPSGHHHHHHHSGLDPSGHHHHHHS--GLDPSGHHHHHHHSGLDPSGHHHHHH 65

QY 51 ---HPHGHHPHCH 60
   | | | | | | | |
Db 66 SGLDPSGHHHHHH 78

```

RESULT 11
US-09-328-352-6551
; Sequence 6551, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6551
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6551

Query Match	15.2%;	Score 147;	DB 4;	Length 339;	
Best Local Similarity	27.8%;	Pred. No. 5.7e-07;			
Matches 47;	Conservative 4;	Mismatches 28;	Indels 90;	Gaps 12;	
QY	3	KGHSHEQH-----	PHGHHPHAH-----	PHEHDTHQHPHG-----	34
Db	132	KGHLHRRHHNGCGHGLDLRLNGHGHHLHRRHHNGCDHGLHNLGHDLHLRRH-HGGCGH			190
QY	35	-----HHPHGHHPHG-----	-----HHPHGHHPHG-----	-----HHPHC	59
Db	191	DFGHLSDRDHHLHRRH-HGGCGHDLRLNDRDHLHRRH-HGGCDHDLGLNDRDHLHR			248
QY	60	HDFQYCGPCDPPPHNQGH-----	-----CCGH-----	GPPPGHLR	91
Db	249	H-----HHGCD---HGLGLSDRDHHLHRRHHHCDCHGDRLSDRDHLR			290

RESULT 12
 US-08-255-457-1
 ; Sequence 1, Application US/08255457
 ; Patent No. 5780040
 ; GENERAL INFORMATION:
 ; APPLICANT: Plaut, Andrew G.
 ; APPLICANT: Gilbert-Rothstein, Joanne V.
 ; APPLICANT: Wright, Andrew
 ; TITLE OF INVENTION: HELICOBACTER PYLORI NICKEL BINDING
 ; TITLE OF INVENTION: PROTEIN
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESSES:
 ; ADDRESSEE: Fish & Richardson
 ; STREET: 225 Franklin Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02110-2804
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/255,457
 ; FILING DATE:
 ; CLASSIFICATION: 530
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Clark, Paul C.
 ; REGISTRATION NUMBER: 30,162
 ; REFERENCE/DOCKET NUMBER: 00398/090001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 542-5070
 ; TELEFAX: (617) 542-8906
 ; TELEX: 200154
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 60 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-255-457-1

	Query Match	15.0%;	Score 144.5;	DB 1;	Length 60;
	Best Local Similarity	40.9%;	Pred. No. 1.6e-07;		
	Matches	27; Conservative	3; Mismatches	25; Indels	11; Gaps
					3;
Qy	19	AHHPEHDTTRQHFGHHHPHGHHPHGHHHPGHCHDFQDYGPCDPPP--HNQG	76		
		:		:	
Dd	2	AHHEEQGGY--HHHHHTHHHHYHGCCEHHHHSSHH-----EEGCCSTSDSHHQEE	52		
Qy	77	HCCHGH	82		
Dd	53	GCCCHG	58		

```

;
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-115-032-1

Query Match      15.0%; Score 144.5; DB 2; Length 60;
Best Local Similarity 40.9%; Pred. No. 1.6e-07;
Matches 27; Conservative 3; Mismatches 25; Indels 11; Gaps 3;

Qy 19 AHHPHEHDTHRQHPHGHHPHGHHPHGHHPHGHHPHCHDQDYGPCDPPP--HNQG 76
    |||:| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2 AHHEEQHGCH--HHHHHTHHHHYHGGEHHHHHHSSHH-----EEGCCSTSDSHHQEE 52
    |||:| | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 77 HCCHGH 82
    |||||
Db 53 GCCHGH 58

RESULT 14
PCT-US95-05772-1
; Sequence 1, Application PC/TUS9505772
; GENERAL INFORMATION:
; APPLICANT: Plaut, Andrew G.
; APPLICANT: Gilbert-Rothstein, Joanne V.
; APPLICANT: Wright, Andrew
; TITLE OF INVENTION: HELICOBACTER PYLORI NICKEL
; TITLE OF INVENTION: BINDING PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street

```



```
;
; FILING DATE: 13-DEC-1991
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Einaudi, Carol P.
; REGISTRATION NUMBER: 32,220
; REFERENCE/DOCKET NUMBER: 02481-1143-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 189 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-152-922A-6

Query Match      14.5% Score 140; DB 1; Length 189;
Best Local Similarity 33.0%; Pred. No. 1.5e-06;
Matches 37; Conservative 5; Mismatches 34; Indels 36; Gaps 11;

QY      3 HKHSHEQHPHGHP-----HAHP-----HEHDTHQHPHGHP-----HGHP-----H 43
Db      7 HAHHAADAH-HAHHAADAHHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHA 64

QY      44 GHHP-----HGHP-----HCHDFOD-YGPCDPPPHNQGHCCHGH 82
Db      65 AHHAADAHHAHAHAADAHHAHAHAADAHHAHAHAADAHHAHAHAADAHHAHAHAADAH 114
```

Search completed: September 23, 2004, 22:56:48
Job time : 27.747 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2004, 22:55:11 ; Search time 101.044 Seconds
(without alignments)
470.989 Million cell updates/sec

Title: US-10-074-225A-5
Perfect score: 966
Sequence: 1 HPHKHSHEQHPGHHPAH.....PSFPLPHKHPLKPDNQFPF 148

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1349238 seqs, 321558718 residues

Total number of hits satisfying chosen parameters: 1349238

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*

2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*

3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*

4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*

5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*

6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*

7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*

8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*

9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*

10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*

11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*

12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*

13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*

14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*

15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*

16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*

17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*

18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	966	100.0	148	14	US-10-074-225A-5
2	966	100.0	525	14	US-10-074-225A-1
3	949	98.2	525	10	US-09-919-039-62
4	543	56.2	526	14	US-10-074-225A-3
5	325	33.6	67	12	US-10-424-599-166051
6	302	31.3	75	9	US-09-730-379B-1
7	280.5	29.0	101	14	US-10-074-225A-6
8	242.5	25.1	378	14	US-10-029-386-33892
9	221	22.9	292	9	US-09-864-761-37944
10	213	22.0	38	9	US-09-730-379B-9
11	213	22.0	38	9	US-09-730-379B-13
12	210	21.7	82	9	US-09-864-761-33313
13	199.5	20.7	130	16	US-10-389-566-2423
14	199.5	20.7	133	16	US-10-389-566-2424
15	199.5	20.7	265	16	US-10-389-566-2422

16	192.5	19.9	342	12	US-10-424-599-281824	Sequence 281824,
17	191.5	19.8	156	9	US-09-864-761-41679	Sequence 41679, A
18	185	19.2	26	9	US-09-809-657-2	Sequence 2, Appli
19	185	19.2	26	9	US-09-848-512A-2	Sequence 2, Appli
20	185	19.2	26	9	US-09-742-494-2	Sequence 2, Appli
21	185	19.2	26	14	US-10-153-312A-4	Sequence 4, Appli
22	185	19.2	78	9	US-09-864-761-37352	Sequence 37352, A
23	182	18.8	253	12	US-10-282-122A-45941	Sequence 45941, A
24	179.5	18.6	87	9	US-09-864-761-33727	Sequence 33727, A
25	179.5	18.6	87	9	US-09-864-761-34744	Sequence 34744, A
26	178.5	18.5	140	16	US-10-389-566-825	Sequence 825, App
27	175	18.1	430	12	US-10-425-114-69999	Sequence 69999, A
28	172	17.8	314	12	US-10-424-599-186228	Sequence 186228,
29	171.5	17.8	144	12	US-10-424-599-160290	Sequence 160290,
30	166	17.2	153	14	US-10-083-357-1014	Sequence 1014, Ap
31	164.5	17.0	695	10	US-09-890-688-130	Sequence 130, App
32	164	17.0	49	9	US-09-864-761-37882	Sequence 37882, A
33	164	17.0	108	14	US-10-029-386-31185	Sequence 31185, A
34	164	17.0	574	14	US-10-156-761-14106	Sequence 14106, A
35	162	16.8	476	15	US-10-360-849A-33	Sequence 33, Appl
36	160.5	16.6	366	12	US-10-406-686A-76	Sequence 76, Appl
37	158	16.4	139	16	US-10-437-963-122803	Sequence 122803,
38	158	16.4	251	14	US-10-029-386-33860	Sequence 33860, A
39	158	16.4	940	12	US-10-221-278-363	Sequence 363, App
40	158	16.4	940	15	US-10-291-172-363	Sequence 363, App
41	157	16.3	2068	16	US-10-437-963-198315	Sequence 198315,
42	156.5	16.2	1130	14	US-10-032-585-7758	Sequence 7758, Ap
43	155.5	16.1	421	14	US-10-156-761-13228	Sequence 13228, A
44	155.5	16.1	865	9	US-09-957-995A-19	Sequence 19, Appl
45	155	16.0	29	9	US-09-730-379B-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-10-074-225A-5
; Sequence 5, Application US/10074225A
; Publication No. US20030082740A1
; GENERAL INFORMATION:
; APPLICANT: DONATE, Fernando
; APPLICANT: PLUNKETT, Marian L
; APPLICANT: HARRIS, Scott
; APPLICANT: MAZAR, Andrew P
; TITLE OF INVENTION: HISTIDINE PROLINE RICH GLYCOPROTEIN (HPRG) AS AN ANTI-ANGIOGENIC
; TITLE OF INVENTION: ANTI-TUMOR AGENT
; FILE REFERENCE: 38342-178463
; CURRENT APPLICATION NUMBER: US/10/074,225A
; CURRENT FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: US 60/268,370
; PRIOR FILING DATE: 2001-02-14
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 148
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-074-225A-5
Query Match 100.0%; Score 966; DB 14; Length 148;
Best Local Similarity 100.0%; Pred. No. 1.4e-69;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 HPHKHSHEQHPGHHPAHHPHEHDTTHRQHPGHHPGHHPGHHPGHHPCH 60
|||||
Db 1 HPHKHSHEQHPGHHPAHHPHEHDTTHRQHPGHHPGHHPGHHPGHHPCH 60
QY 61 DFQDYGFCDDPPPHNQHCCHGHPGPPGHLRRRPGKGPFPFHCRCQIGSVYRLPPLRKGEV 120
|||||
Db 61 DFQDYGFCDDPPPHNQHCCHGHPGPPGHLRRRPGKGPFPFHCRCQIGSVYRLPPLRKGEV 120
QY 121 LPLPEANFPSPFLPHKHPLKPDNQFPF 148
|||||

D^b 121 LPLPEANFSPFLPHHKHPLKPDNQFPF 148

RESULT 2

US-10-074-225A-1
; Sequence 1, Application US/10074225A
; Publication No. US20030082740A1
; GENERAL INFORMATION:
; APPLICANT: DONATE, Fernando
; APPLICANT: PLUNKETT, Marian L
; APPLICANT: HARRIS, Scott
; APPLICANT: MAZAR, Andrew P
; TITLE OF INVENTION: HISTIDINE PROLINE RICH GLYCOPROTEIN (HPRG) AS AN ANTI-ANGIOGENIC
; TITLE OF INVENTION: ANTI-TUMOR AGENT
; FILE REFERENCE: 38342-178463
; CURRENT APPLICATION NUMBER: US/10/074, 225A
; CURRENT FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: US 60/268,370
; PRIOR FILING DATE: 2001-02-14
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-074-225A-1

```
Query Match      100.0%; Score 966; DB 14; Length 525;
Best Local Similarity 100.0%; Pred. No. 4.4e-69;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy

1 HPHKHSHEQHPRGHHPRFAHHPHEHTDTHRQHPHGHHPRGHHPHGHHPRGHHPHGHHPRCH 60

350 HPHKHSHEQHPRGHHPRFAHHPHEHTDTHRQHPHGHHPRGHHPHGHHPRGHHPHGHHPRCH 40

Dd

QY 61 DFQYGPDPDPHNOGHCCHGGPPCHLRRRGKGRPFHCQIGSVYRLPLRKGEV 120
410 DFQYGPDPDPHNOGHCCHGGPPCHLRRRGKGRPFHCQIGSVYRLPLRKGEV 469

QY 121 LPLPEANFSPFPLPHHKHPLKPDNQPF 148
|||
DQ 470 LPLPEANFSPFPLPHHKHPLKPDNQPF 497

RESULT 3

```

US-09-919-039-62
; Sequence 62, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 62
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030108871A1 085596CD1
US-09-919-039-62

```

```
Query Match      98.2%; Score 949; DB 10; Length 525;
Best Local Similarity 98.6%; Pred. NO. 9.9e-68;
Matches 146; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

Qy 1 HPHKHSHSQHPHGHPHHAHPHPEHDTTHRQHPHGHPHGHHPHGHHPHGHHPHGHHPHGH 60

Db	350	HPHKHSHEQHPHGHHPPHAAHHPHEHDTHQHPHGHHPPHGHHPPHGHHPPHGHHPPHCH	409
QY	61	DFQDYGCDDPPPHNQHCCHGCGPPPGHLRRRGGPGKGPFFHCRQIGSVYRLPPLRKGEV	1200
Db	410	DFQDYGCDDPPPHNQHCCHGCGPPPGHLRRRGGPGKGPFFHCRQIGSVYRLPPLRKGEV	4699
QY	121	LPLPEANFPSPPLPHKHPLKPDNQPPF	148
Db	470	LPLPEANFPSPPLPHKHPLKPDNQPPF	497

RESULT 4

US-10-074-225A-3
; Sequence 3, Application US/10074225A
; Publication No. US20030082740A1
; GENERAL INFORMATION:
; APPLICANT: DONATE, Fernando
; APPLICANT: PLUNKETT, Marian L
; APPLICANT: HARRIS, Scott
; APPLICANT: MAZAR, Andrew P
; TITLE OF INVENTION: HISTIDINE PROLINE RICH GLYCOPROTEIN (HPRG) AS AN ANTI-ANGIOGENIC
; TITLE OF INVENTION: ANTI-TUMOR AGENT
; FILE REFERENCE: 38342-178463
; CURRENT APPLICATION NUMBER: US/10/074,225A
; CURRENT FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: US 60/268,370
; PRIOR FILING DATE: 2001-02-14
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 526
; TYPE: PRT
; ORGANISM: Lepus americanus
US-10-074-225A-3

Query Match	56.2%;	Score 543;	DB 14;	Length 526;
Best Local Similarity	57.6%;	Pred. NO. 2.1e-35;		
Matches 98;	Conservative	8;	Mismatches 40;	Indels 24;
				Gaps 4;

	QY	2	RKCHSHEQHGHCHPHNA-----HH-----PHEHDTHQHPHGHNHPGHNRP	42
	dD	330	PNGNHRPGPPRHGHNRPHGGPRRPHRGPPRRHNPARGPPRRHGPPRHGEPPRHGNRPFGPPP	389

QY 43 HGHHPHGHHPHCHDFQDYGPCDPHPHNG---HCCHGHGPPGHHLRRRGGPKGP 98
 390 HGHHPHGHHPHGHGHGHGHGFDHGGPCDPHSHKEGPODLH-OHAMGPPXHPKRGKRGPKGH 44

	99	RPFHCRIGSVYRLPPLRKGEVLPLPEANFPSPFLPHHKHPLKPDNQPPF	148
QY		: : : : : :	
	449	FDFHWERTGSVVOI PPIOKGEVI PL PEANFPOLLRNTHHTPL KPEIOEPF	498
N6		: : : : : :	

RESULT 5

```

US-10-424-599-166051
; Sequence 166051, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 166051
; LENGTH: 67
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure

```

```

; LOCATION: (1)..(67)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_120960C.1.pap
US-10-424-599-166051

Query Match      33.6%; Score 325; DB 12; Length 67;
Best Local Similarity 92.2%; Pred. No. 7.4e-19;
Matches 47; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 HPHKHSHEQHPGHHPHAHPHEHDTTHRQHPGHHPGHHPGHHPGHHPGH 51
    |||||
Db 17 HPHKHSHEQHPGHHPHAHPHEHDTTHRQHPGHHPGHHPGHHPGHHPGH 67
    |||||

RESULT 6
US-09-730-379B-1
; Sequence 1, Application US/09730379B
; Patent No. US20010041670A1
; GENERAL INFORMATION:
; APPLICANT: Simantov M.D., Ronit
; APPLICANT: Silverstein M.D., Roy L.
; TITLE OF INVENTION: THROMBOSPONDIN-BINDING REGION OF HISTIDINE-RICH
; FILE REFERENCE: 955-7P/CON
; CURRENT APPLICATION NUMBER: US/09/730,379B
; CURRENT FILING DATE: 2000-12-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 1
; LENGTH: 75
; TYPE: PRT
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION: PUBLICATION INFORMATION
; JOURNAL: Biochemistry
; VOLUME: 25
; ISSUE: 8
; PAGES: 2220-2225
; DATE: 1986
; DATABASE ACCESSION NUMBER: GenBank/P04196
US-09-730-379B-1

Query Match      31.3%; Score 302; DB 9; Length 75;
Best Local Similarity 100.0%; Pred. No. 5.6e-17;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 GPRFHCRIQSVYRLPLRKGEVLPPLPEANFSPPLPHKHPKLPDNPQPP 148
    |||||
Db 1 GPRFHCRIQSVYRLPLRKGEVLPPLPEANFSPPLPHKHPKLPDNPQPP 52
    |||||

RESULT 7
US-10-074-225A-6
; Sequence 6, Application US/10074225A
; Publication No. US20030082740A1
; GENERAL INFORMATION:
; APPLICANT: DONATE, Fernando
; APPLICANT: PLUNKETT, Marian L
; APPLICANT: HARRIS, Scott
; APPLICANT: MAZAR, Andrew P
; TITLE OF INVENTION: HISTIDINE PROLINE RICH GLYCOPROTEIN (HPRG) AS AN ANTI-ANGIOGENIC
; FILE REFERENCE: 38342-178463
; CURRENT APPLICATION NUMBER: US/10/074,225A
; CURRENT FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: US 60/268,370
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Lepus americanus
```

```

US-10-074-225A-6

Query Match      29.0%; Score 280.5; DB 14; Length 101;
Best Local Similarity 51.6%; Pred. No. 3.8e-15;
Matches 47; Conservative 1; Mismatches 24; Indels 19; Gaps 2;

QY 2 PHKHSHEQHPGHHPHA-----HH-----PHEHDTTHRQHPGHHPGHHP 42
    |||||
Db 10 PHGHHPGHPPPHGHHPGHPPPHGHHPGHPPPHGHHPGHPPPHGHHPGH 69
    |||||
QY 43 HGHPHPGHHPGHHPGHCHDFQDYGPCDPPPH 73
    |||||
Db 70 HGHPHPGHPPPHGHHPGHCHDFQDYGPCDPPSH 100
    |||||

RESULT 8
US-10-029-386-33892
; Sequence 33892, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33892
; LENGTH: 378
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP001751.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.6
US-10-029-386-33892

Query Match      25.1%; Score 242.5; DB 14; Length 378;
Best Local Similarity 47.1%; Pred. No. 1.4e-11;
Matches 41; Conservative 1; Mismatches 34; Indels 11; Gaps 2;

QY 1 HPHKHSHEQHPGHHPHAHPHEHDTTHRQHPGHHPGHHPGHHPGHHPGH 55
    |||||
Db 39 HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH 98
    |||||
QY 56 HPHCHDFQDYGPCDPPPHNQGHCHGH 82
    |||||
Db 99 HHHHHHHHHY-----HHHQHHHHHHH 119
    |||||

RESULT 9
US-09-864-761-37944
; Sequence 37944, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
```



```

RESULT 15
US-10-389-566-2422
; Sequence 2422, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2422
; LENGTH: 265
; TYPE: PRT
; ORGANISM: Bacillus anthracis
US-10-389-566-2422

```

```
Query Match      20.7%; Score 199.5; DB 16; Length 265;  
Best Local Similarity 44.6%; Pred. No. 2.6e-08;  
Matches 45; Conservative 2; Mismatches 35; Indels 19; Gaps 8;  
  
QY    1 HPHKHSHEQHPGCHHPPHAHHPPHEHTHRQHPPH-GHHPPH-GHH-PHGHHP----- 52  
Dð    111 HGHCCHHHQCHRGHHGGHHHHHQGHNGHHGHHHQGHNGHHQGHNGHHGHHGHHGHHQQQVH 170  
  
QY    53 -HGHHPHCHD----FDYGPCDPppPHNQGCCHGCGPPPGH 88  
Dð    171 HHGHH-HIHPQAIIYQTH-----QGHCQHHPDHGHHGHGQH 205
```

Search completed: September 23, 2004, 23:11:16
Job time : 103.044 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2004, 21:23:24 ; Search time 68.9558 Seconds
(without alignments)
413.849 Million cell updates/sec

Title: US-10-074-225A-6
Perfect score: 697
Sequence: 1 SVNIIHRPPPHGHHPGPPPP.....HPPHGHGPHDHGPCDPPSHK 101

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	697	100.0	526	5	ABB79805 Rabbit hi
2	314	45.1	525	5	ABB79804 Human his
3	314	45.1	525	8	ADE76897 Human pro
4	280.5	40.2	148	5	ABB79807 Rabbit hi
5	280.5	40.2	148	5	ABB79806 Human his
6	264.5	37.9	491	4	ABG09949 Novel hum
7	259.5	37.2	296	7	ADD45442 Rat prote
8	259.5	37.2	296	7	ADE57105 Rat prote
9	246	35.3	274	7	ADD47240 Rat prote
10	246	35.3	274	7	ADD48703 Rat prote
11	233.5	33.5	274	2	AAR04119 ORF2 of E
12	233.5	33.5	309	2	AAR04116 ORF1 of E
13	233.5	33.5	331	7	ADD45740 Human Pro
14	233.5	33.5	424	6	ABO52997 Human spl
15	231.5	33.2	192	6	AAO30246 Human Sap
16	231.5	33.2	208	6	AAO30188 Human cp3
17	231	33.1	267	7	AAO30408 Human sec
18	229.5	32.9	124	2	AAW50192 Amino aci
19	228.5	32.8	93	4	AAW71667 Human col
20	228.5	32.8	93	4	AAW71654 Human col
21	228.5	32.8	102	4	ABB68218 Drosophil
22	228	32.7	247	6	ABU11891 Human ABC
23	225	32.3	80	4	ABB61183 Drosophil
24	223	32.0	594	4	ABB61362 Drosophil
25	223	32.0	1130	5	ABP73921 Candida a

26	221	31.7	1378	5	ABB57176	Abb57176 Mouse isc
27	215.5	30.9	104	5	ABB78538	Abb78538 Ser-Pro-P
28	215	30.8	162	7	AAO30407	Aao30407 Human sec
29	210	30.1	487	2	AAR22380	Aar22380 Antigen m
30	209.5	30.1	205	5	ABR39650	Abr39650 A. thalia
31	208	29.8	865	3	AAB26153	Aab26153 Neospora
32	205	29.4	815	4	ABB70499	Abb70499 Drosophil
33	205	29.4	815	6	ABP54340	Abp54340 Drosophil
34	198.5	28.5	147	5	ABP43151	Abp43151 Human ova
35	198	28.4	164	7	ADD47283	Add47283 Rat Prote
36	196	28.1	82	2	AAW50193	Aaw50193 Amino aci
37	195.5	28.0	99	4	AAO02076	Aao02076 Human pol
38	195.5	28.0	523	7	ADC87177	Adc87177 Human GPC
39	194.5	27.9	926	4	ABB65135	Abb65135 Drosophil
40	193	27.7	749	4	ABB62915	Abb62915 Drosophil
41	192.5	27.6	455	2	AAR12362	Aar12362 Octopus r
42	192.5	27.6	940	4	AAU28194	Aau28194 Novel hum
43	192	27.5	202	7	ADD48701	Add48701 Rat Prote
44	192	27.5	351	2	AAR24393	Aar24393 Sequence
45	192	27.5	464	6	ABO52933	Abo52933 Human spl

ALIGNMENTS

RESULT 1

ABB79805
ID ABB79805 standard; protein; 526 AA.

XX ABB79805;

XX 25-NOV-2002 (first entry)

DT Rabbit histidine proline rich glycoprotein.

DE Histidine proline rich glycoprotein; HPRG; rabbit; antiangiogenic;

XX cytotostatic; antiarteriosclerotic; antiinflammatory; antidiabetic;

KW gynaecological; antiarthritic; antiulcer; osteopathic; antitumour;

KW ophthalmological; nootropic; neuroprotective; antiparkinsonian; chelator.

OS Oryctolagus cuniculus.

XX Key Location/Qualifiers

FT Domain 251..296

FT Domain /note= "Proline-rich domain"

FT Domain 321..421

FT Domain /note= "Histidine-proline-rich domain, region also

FT Domain specifically claimed in Claim 1"

XX WO200264621-A2.

XX 22-AUG-2002.

XX 14-FEB-2002; 2002WO-US004336.

XX 14-FEB-2001; 2001US-0268370P.

XX (ATTE-) ATTENUON LLC.

XX Donate F, Harris S, Plunkett ML, Mazar AP;

XX WPI; 2002-666989/71.

XX P-PSDB; ABN84911.

XX New histidine-proline rich glycoprotein (HPRG) polypeptide, useful for

PT diagnosing or treating diseases associated with undesired cell migration,

PT invasion, proliferation, or angiogenesis, e.g. cancer, atherosclerosis.

XX Claim 1; Page 12-13; 82pp; English.

PS The present sequence is the protein sequence of rabbit histidine proline

XX rich glycoprotein (HPRG), a proteinaceous chelator that can be used to

CC inhibit angiogenesis and treat cancer. Claimed anti-angiogenic

CC polypeptides or peptides comprise: the histidine-proline-rich (H/P) domain (see ABB79806) of human HPRG; an H/P domain of rabbit HPRG (see ABB79807); a variant of these that is capable of inhibiting angiogenesis, endothelial cell proliferation or endothelial tube formation in vitro or in vivo; or a pentapeptide having the generic sequence given in ABB79808. Also claimed are: chemically synthesised or recombinantly produced peptide multimers; a diagnostically or therapeutically labeled anti-angiogenic polypeptide, peptide or peptide multimer; a diagnostically useful HPRG-related composition, comprising the diagnostically labeled polypeptide, peptide or peptide multimer and a carrier; an antibody specific for an epitope of HPRG that is present in the H/P domain of human or rabbit HPRG, and which binds to HPRG or to any of the domains in a way which inhibits the anti-angiogenic activity of HPRG or the domain, or an antigen-binding fragment of the antibody; a method for inhibiting cell migration, cell invasion, cell proliferation or angiogenesis, or for inducing apoptosis; a method for treating a subject having a disease or condition associated with undesired cell migration, invasion, proliferation, or angiogenesis; a method of stimulating or inhibiting angiogenesis in a subject; a method of detecting the presence of HPRG or its cleavage product or its peptide in a biological sample; isolated nucleic acids encoding the polypeptide, peptide or peptide multimer; an expression vector; transformed or transfected cells; a method of providing to a cell, tissue or organ an angiogenesis-inhibitory amount of HPRG, an H/P domain of HPRG or its pentapeptide, or the peptide multimer; an affinity ligand useful for binding to, or isolating, an HPRG-binding molecule or cells expressing the binding molecule, comprising the polypeptide, peptide or peptide multimer, immobilised to a solid support or carrier; and a method of isolating HPRG-binding molecule, or isolating or enriching cells expressing HPRG-binding site or receptor, from a complex mixture. The compositions and methods are useful in diagnosing or treating a disease or condition associated with undesired cell migration, invasion, proliferation, or angiogenesis, such as cancer, atherosclerosis, diabetic retinopathy, inflammation, endometriosis, arthritis, peptic ulcers, or fractures. HPRG is especially useful in inhibiting the growth of primary tumours or metastases, and may also be used in treating neurodegenerative diseases like Alzheimer's disease, Parkinson's disease and amyotrophic lateral sclerosis. The antibodies are stimulators of angiogenesis and are useful for promoting neovascularization in pertinent disease states, and in various immunoassays

SQ Sequence 526 AA;

Query Match 100.0%; Score 697; DB 5; Length 526;
Best Local Similarity 100.0%; Pred. No. 5.5e-51;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVNIHRPPPHGHPGPPPHGHHGHPGPPPHGPPPHGPPPHGPPPHGPPPH 60
Db |||||
321 SVNIHRPPPHGHPGPPPHGHHGHPGPPPHGPPPHGPPPHGPPPHGPPPH 380
QY 61 GHPPHGGPPPHGHPGPPPHGHPGHPGHPGHPGHPGHPGHPGHPGHPGHP 101
Db |||||
381 GHPPHGGPPPHGHPGPPPHGHPGHPGHPGHPGHPGHPGHPGHPGHPGHP 421

RESULT 2

ABB79804

ID ABB79804 standard; protein; 525 AA.

XX

AC ABB79804;

XX

DT 25-NOV-2002 (first entry)

XX

DE Human histidine proline rich glycoprotein.

XX

KW Histidine proline rich glycoprotein; HPRG; human; antiangiogenic; cytostatic; antiarteriosclerotic; antiinflammatory; antidiabetic; gynaecological; antiarthritic; antiulcer; osteopathic; antitumour; ophthalmological; nootropic; neuroprotective; antiparkinsonian; chelator.

XX

OS Homo sapiens.

XX

SQ Sequence 525 AA;

FH Key Location/Qualifiers
FT Domain 276..321
FT /note= "Proline-rich domain"
FT Domain 350..497
FT /note= "Histidine-proline-rich domain, region also specifically claimed in Claim 1"
XX WO200264621-A2.
PN 22-AUG-2002.
PD 14-FEB-2002; 2002WO-US004336.
XX 14-FEB-2001; 2001US-0268370P.
PR (ATTE-) ATTENUON LLC.
XX Donate F, Harris S, Plunkett ML, Mazar AP;
PI WPI; 2002-666989/71.
XX P-PSDB; ABN84910.
DR
XX New histidine-proline rich glycoprotein (HPRG) polypeptide, useful for diagnosing or treating diseases associated with undesired cell migration, invasion, proliferation, or angiogenesis, e.g. cancer, atherosclerosis.
PS Claim 1; Page 11; 82pp; English.
XX
CC The present sequence is the protein sequence of human histidine proline rich glycoprotein (HPRG), a proteinaceous metal chelator that can be used to inhibit angiogenesis and treat cancer. Claimed anti-angiogenic polypeptides or peptides comprise: the histidine-proline-rich (H/P) domain (see ABB79806) of human HPRG; an H/P domain of rabbit HPRG (see ABB79807); a variant of these that is capable of inhibiting angiogenesis, endothelial cell proliferation or endothelial tube formation in vitro or in vivo; or a pentapeptide having the generic sequence given in ABB79808. Also claimed are: chemically synthesised or recombinantly produced peptide multimers; a diagnostically or therapeutically labeled anti-angiogenic polypeptide, peptide or peptide multimer; a diagnostically useful HPRG-related composition, comprising the diagnostically labeled polypeptide, peptide or peptide multimer and a carrier; an antibody specific for an epitope of HPRG that is present in the H/P domain of human or rabbit HPRG, and which binds to HPRG or to any of the domains in a way which inhibits the anti-angiogenic activity of HPRG or the domain, or an antigen-binding fragment of the antibody; a method for inhibiting cell migration, cell invasion, cell proliferation or angiogenesis, or for inducing apoptosis; a method for treating a subject having a disease or condition associated with undesired cell migration, invasion, proliferation, or angiogenesis; a method of stimulating or inhibiting angiogenesis in a subject; a method of detecting the presence of HPRG or its cleavage product or its peptide in a biological sample; isolated nucleic acids encoding the polypeptide, peptide or peptide multimer; an expression vector; transformed or transfected cells; a method of providing to a cell, tissue or organ an angiogenesis-inhibitory amount of HPRG, an H/P domain of HPRG or its pentapeptide, or the peptide multimer; an affinity ligand useful for binding to, or isolating, an HPRG-binding molecule or cells expressing the binding molecule, comprising the polypeptide, peptide or peptide multimer, immobilised to a solid support or carrier; and a method of isolating HPRG-binding site or receptor, from a complex mixture. The compositions and methods are useful in diagnosing or treating a disease or condition associated with undesired cell migration, invasion, proliferation, or angiogenesis, such as cancer, atherosclerosis, diabetic retinopathy, inflammation, endometriosis, arthritis, peptic ulcers, or fractures. HPRG is especially useful in inhibiting the growth of primary tumours or metastases, and may also be used in treating neurodegenerative diseases like Alzheimer's disease, Parkinson's disease, and amyotrophic lateral sclerosis. The antibodies are stimulators of angiogenesis and are useful for promoting neovascularization in pertinent disease states, and in various immunoassays

CC of the domains in a way which inhibits the anti-angiogenic activity of
CC HPRG or the domain, or an antigen-binding fragment of the antibody; a
CC method for inhibiting cell migration, cell invasion, cell proliferation
CC or angiogenesis, or for inducing apoptosis; a method for treating a
CC subject having a disease or condition associated with undesired cell
CC migration, invasion, proliferation, or angiogenesis; a method of
CC stimulating or inhibiting angiogenesis in a subject; a method of
CC detecting the presence of HPRG or its cleavage product or its peptide in
CC a biological sample; isolated nucleic acids encoding the polypeptide,
CC peptide or peptide multimer; an expression vector; transformed or
CC transfected cells; a method of providing to a cell, tissue or organ an
CC angiogenesis-inhibitory amount of HPRG, an H/P domain of HPRG or its
CC pentapeptide, or the peptide multimer; an affinity ligand useful for
CC binding to, or isolating, an HPRG-binding molecule or cells expressing
CC the binding molecule, comprising the polypeptide, peptide or peptide
CC multimer, immobilised to a solid support or carrier; and a method of
CC isolating HPRG-binding molecule, or isolating or enriching cells
CC expressing HPRG-binding site or receptor, from a complex mixture. The
CC compositions and methods are useful in diagnosing or treating a disease
CC or condition associated with undesired cell migration, invasion,
CC proliferation, or angiogenesis, such as cancer, atherosclerosis, diabetic
CC retinopathy, inflammation, endometriosis, arthritis, peptic ulcers, or
CC fractures. HPRG is especially useful in inhibiting the growth of primary
CC tumours or metastases, and may also be used in treating neurodegenerative
CC diseases like Alzheimer's or Parkinson's disease. The antibodies are
CC stimulators of angiogenesis and are useful for promoting
CC neovascularization in pertinent disease states, and in various
CC immunoassays

XX Sequence 148 AA;

Query Match 40.2%; Score 280.5; DB 5; Length 148;
Best Local Similarity 51.6%; Pred. No. 1.7e-16;
Matches 47; Conservative 1; Mismatches 24; Indels 19; Gaps 2;

QY 10 PHGHPHGGPPPHGHHPHGHPPHGHPPHPPHPPHPPHGHPPHGHPPHGHPPH 69
Db 2 PKHHSHEQHHPHGHHPH-----HH-----PHEHDTTHRQHPHGHHPH 42
QY 70 HGHPHGGPPPHGHHPHGHGHGHGHGHGHGHGHGHGHGHGHGHGHGHGHGH 100
Db 43 HGHPHGHHPHGHHPHGHCHDFQDYGPCDPPPH 73

RESULT 5
ABB79806

ID ABB79806 standard; protein; 148 AA.

XX ABB79806;

XX 25-NOV-2002 (first entry)

XX Human histidine proline rich glycoprotein His/Pro-rich domain.

XX Histidine proline rich glycoprotein; HPRG; human; antiangiogenic;
XX cystostatic; antiarteriosclerotic; antiinflammatory; antidiabetic;
XX gynaecological; antiarthritic; antiulcer; osteopathic; antitumour;
XX ophthalmological; nootropic; neuroprotective; antiparkinsonian.

OS Homo sapiens.

XX WO200264621-A2.

XX 22-AUG-2002.

XX 14-FEB-2002; 2002WO-US004336.

XX 14-FEB-2001; 2001US-0268370P.

XX (ATTE-) ATTENUON LLC.

PI Donate F, Harris S, Plunkett ML, Mazar AP;

DR

DR WPI; 2002-666989/71.

XX P-PSDB; ABN84910.

XX New histidine-proline rich glycoprotein (HPRG) polypeptide, useful for
PT diagnosing or treating diseases associated with undesired cell migration,
PT invasion, proliferation, or angiogenesis, e.g. cancer, atherosclerosis.

XX Claim 1; Page 13; 82pp; English.

XX The present sequence is the protein sequence of the histidine-proline-
rich (H/P) domain of human anti-angiogenic histidine proline rich
glycoprotein (HPRG, see also ABB79804). Claimed anti-angiogenic
polypeptides or peptides comprise: the H/P domain of human HPRG; an H/P
domain of rabbit HPRG (see ABB12345); a variant of these that is capable
of inhibiting angiogenesis, endothelial cell proliferation or endothelial
tube formation in vitro or in vivo; or a pentapeptide having the generic
sequence given in ABB79808. Also claimed are: chemically synthesised or
recombinantly produced peptide multimers; a diagnostically or
therapeutically labeled anti-angiogenic polypeptide, peptide or peptide
multimer; a diagnostically useful HPRG-related composition, comprising
the diagnostically labeled polypeptide, peptide or peptide multimer and a
carrier; an antibody specific for an epitope of HPRG that is present in
the H/P domain of human or rabbit HPRG, and which binds to HPRG or to any
of the domains in a way which inhibits the anti-angiogenic activity of
HPRG or the domain, or an antigen-binding fragment of the antibody; a
method for inhibiting cell migration, cell invasion, cell proliferation
or angiogenesis, or for inducing apoptosis; a method for treating a
subject having a disease or condition associated with undesired cell
migration, invasion, proliferation, or angiogenesis; a method of
stimulating or inhibiting angiogenesis in a subject; a method of
detecting the presence of HPRG or its cleavage product or its peptide in
a biological sample; isolated nucleic acids encoding the polypeptide,
peptide or peptide multimer; an expression vector; transformed or
transfected cells; a method of providing to a cell, tissue or organ an
angiogenesis-inhibitory amount of HPRG, an H/P domain of HPRG or its
pentapeptide, or the peptide multimer; an affinity ligand useful for
binding to, or isolating, an HPRG-binding molecule or cells expressing
the binding molecule, comprising the polypeptide, peptide or peptide
multimer, immobilised to a solid support or carrier; and a method of
isolating HPRG-binding molecule, or isolating or enriching cells
expressing HPRG-binding site or receptor, from a complex mixture. The
compositions and methods are useful in diagnosing or treating a disease
or condition associated with undesired cell migration, invasion,
proliferation, or angiogenesis, such as cancer, atherosclerosis, diabetic
retinopathy, inflammation, endometriosis, arthritis, peptic ulcers, or
fractures. HPRG is especially useful in inhibiting the growth of primary
tumours or metastases; and may also be used in treating neurodegenerative
diseases like Alzheimer's or Parkinson's disease. The antibodies are
stimulators of angiogenesis and are useful for promoting
neovascularization in pertinent disease states, and in various
immunoassays

XX Sequence 148 AA;

Query Match 40.2%; Score 280.5; DB 5; Length 148;
Best Local Similarity 51.6%; Pred. No. 1.7e-16;
Matches 47; Conservative 1; Mismatches 24; Indels 19; Gaps 2;

QY 10 PHGHPHGGPPPHGHHPHGHPPHGHPPHPPHPPHGHPPHGHPPHGHPPH 69
Db 2 PKHHSHEQHHPHGHHPH-----HH-----PHEHDTTHRQHPHGHHPH 42

QY 70 HGHPHGGPPPHGHHPHGHGHGHGHGHGHGHGHGHGHGHGHGHGHGHGHGH 100

Db 43 HGHPHGHHPHGHHPHGHCHDFQDYGPCDPPPH 73

RESULT 6

ABG09949

ID ABG09949 standard; protein; 491 AA.

XX ABG09949;

XX

DT 13-FEB-2002 (first entry)
XX Novel human diagnostic protein #9940.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
KW Homo sapiens.
XX WO200175067-A2.
OS 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US008631.
PN 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
PI N-PSDB; AAS74136.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 20; SEQ ID NO 40308; 103pp; English.
PS The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activities in
CC polypeptide and polynucleotide sequences have application of mutations
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 491 AA;
Query Match 37.9%; Score 264.5; DB 4; Length 491;
Best Local Similarity 41.8%; Pred. No. 1e-14;
Matches 56; Conservative 2; Mismatches 35; Indels 41; Gaps 8;
QY 6 HRPPP-----HGHPHG-----PPPHGHPH-GPPPHGHPHGGPPPHPPH----- 45
Db 136 HRPPPAQVTTSTGHHQHGHQHRPPPAHGHHQHSSPPAHGHHQHRPPAHGHHQHHS 195
QY 46 GPPPHGHPHGGPPPHGHPHGGPPH-----PHGHPHGGPPH-----HGHPHG 86
Db 196 SPPAHGHHQHGHQHRPPPAHGHHQHSSPPAHGHHQHGHQHRPPAHGHHQHSSPYQ 255
QY 87 HGFDHGFCDDPPSH 100
Db 256 HGHHQHHRP--PPAH 267

RESULT 7
ADD45442
ID ADD45442 standard; protein; 296 AA.
XX
AC ADD45442;
XX 29-JAN-2004 (first entry)
DT
XX Rat Protein L17318, SEQ ID NO 10875.
DE
XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX
OS Rattus norvegicus.
XX
PN WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
PA (GEHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX
PI Woolf C, D'urso D, Befort K, Costigan M;
XX
DR WPI; 2003-268312/26.
DR GENBANK; L17318.
XX
XX New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
PT
XX
PS Claim 1; Page; 1017pp; English.
XX
CC The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a rat protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 296 AA;
Query Match 37.2%; Score 259.5; DB 7; Length 296;
Best Local Similarity 38.4%; Pred. No. 1.7e-14;
Matches 63; Conservative 1; Mismatches 27; Indels 73; Gaps 10;
QY 2 VNIHRPPPHG---HHPH-----GPPPHG-----HHPHGGPPPHGHP-- 34

Db 81 VNRPERPPQHGNGHHHPHPPAAGPQRPQPGSPQGGPPPPGQQRPPQGGPQQR 140
QY 35 -----PHGPPP-----RHPPHGPPPHGHP-----PHGPPPHGHP-----PHGPPPHG 71
Db 141 PPQGSPPQGGPPPPGQQRPPQGGPQRPQPGSPQGGPQQRPPQGGPQQRAPQGGPP 200

QY 72 HP-----PHGP-----PPHGHPHGFGHDHGCDPP 98
Db 201 GPQRPQPGSPQGGPPPPGQQRPPQGGPPQ-----GPQRP 238

RESULT 8
ADE57105
ID ADE57105 standard; protein; 296 AA.
XX
AC ADE57105;
DT 29-JAN-2004 (first entry)
XX
DE Rat Protein L17318, SEQ ID NO 2965.
KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX
OS Rattus norvegicus.
XX
PN WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
PA (GEHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX
PI Woolf C, D'urso D, Befort K, Costigan M;
XX
DR WPI; 2003-268312/26.
DR GENBANK; L17318.
XX

PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
PS Claim 1; Page; 1017pp; English.
XX
CC The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a rat protein (shown in Table 2 of

CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 296 AA;

Query Match 37.2%; Score 259.5; DB 7; Length 296;
Best Local Similarity 38.4%; Pred. No. 1.7e-14;
Matches 63; Conservative 1; Mismatches 27; Indels 73; Gaps 10;

QY 2 VNIHPPPHG---HHPH-----GPPPHG-----HHPHGPPPHGHP-- 34
Db 81 VNRPERPPQHGNGHHHPHPPAAGPQRPQPGSPQGGPPPPGQQRPPQGGPQQR 140
QY 35 -----PHGPPP-----RHPPHGPPPHGHP-----PHGPPPHGHP-----PHGPPPHG 71
Db 141 PPQGSPPQGGPPPPGQQRPPQGGPQRPQPGSPQGGPQQRPPQGGPQQRAPQGGPP 200
QY 72 HP-----PHGP-----PPHGHPHGFGHDHGCDPP 98
Db 201 GPQRPQPGSPQGGPPPPGQQRPPQGGPPQ-----GPQRP 238

RESULT 9
ADD47240
ID ADD47240 standard; protein; 274 AA.
XX
AC ADD47240;
DT 29-JAN-2004 (first entry)
XX
DE Rat Protein AAA42064, SEQ ID NO 12934.
XX
KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX
OS Rattus norvegicus.
XX
PN WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
PA (GEHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX
PI Woolf C, D'urso D, Befort K, Costigan M;
XX
DR WPI; 2003-268312/26.
DR GENBANK; AAA42064.
XX

PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
PS Claim 1; Page; 1017pp; English.
XX
CC The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially

CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC injury (CCI) and spared nerve injury (SNI) in an animal (e.g. gene
CC therapy). The sequence presented is a rat protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 274 AA;

Query Match 35.3%; Score 246; DB 7; Length 274;
Best Local Similarity 47.1%; Pred. No. 2.2e-13;
Matches 57; Conservative 1; Mismatches 29; Indels 34; Gaps 9;
QY 8 PPHG--HHPHGPPPHG---HHPHGPPPHGHGHP---PHGPPP-----RHPPHGPPPHGHGHP- 53
Db 104 PPGGPPQRPQGPPPPGGPQGGPPPGPQGGPPPGPQGGPPPGPQGGPPPPGGPQ 163
QY 54 --PHGPPPHG---HPPHGPPPHGHP---PHGPPPHGHPHGHGFHDHGP-----CDP 97
Db 164 QGPGPPPPGGPQSPPPGPPPGPQGGPPPGPQGGPPPPGGPQGGP-----QGPPPPGGPQQDP 219
QY 98 P 98
Db 220 P 220

RESULT 10
ADD48703
ID ADD48703 standard; protein; 274 AA.
XX
AC ADD48703;
XX 29-JAN-2004 (first entry)
XX Rat Protein AAA42064, SEQ ID NO 14412.
DE
XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX
OS Rattus norvegicus.
XX
XX WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
XX 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
XX (GEHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX
PI Woolf C, D'urso D, Befort K, Costigan M;
XX
DR WPI; 2003-268312/26.
DR GENBANK; AAA42064.
XX
PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
PS Claim 1; Page; 1017pp; English.

XX The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a rat protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 274 AA;

Query Match 35.3%; Score 246; DB 7; Length 274;
Best Local Similarity 47.1%; Pred. No. 2.2e-13;
Matches 57; Conservative 1; Mismatches 29; Indels 34; Gaps 9;
QY 8 PPHG--HHPHGPPPHG---HHPHGPPPHGHGHP---PHGPPP-----RHPPHGPPPHGHGHP- 53
Db 104 PPGGPPQRPQGPPPPGGPQGGPPPGPQGGPPPGPQGGPPPPGGPQSPPPGGPQ 163
QY 54 --PHGPPPHG---HPPHGPPPHGHP---PHGPPPHGHPHGHGFHDHGP-----CDP 97
Db 164 QGPGPPPPGGPQSPPPGPPPGPQGGPPPGPQGGPPPPGGPQGGP-----QGPPPPGGPQQDP 219
QY 98 P 98
Db 220 P 220

RESULT 11
AAR04119
ID AAR04119 standard; protein; 274 AA.
XX
AC AAR04119;
XX
XX 24-OCT-2003 (revised)
DT 25-MAR-2003 (revised)
DT 05-JUN-1990 (first entry)
XX
DE ORF2 of Enod2b genomic clone.
XX
XX ORF2; Enod2; soybean; root nodule; nodulin 75.
OS Glycine max; (L) Merr. cv. Williams.
XX
PN EP349338-A.
XX
PD 03-JAN-1990.
XX
XX 01-JUL-1988; 88US-00214297.
PF
XX 01-JUL-1988; 88US-00214297.
PR
XX (LUBR) LUBRIZOL GENETICS INC.

polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (Chung), chronic constriction injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene therapy). The sequence presented is a human protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 331 AA;

Query Match 33.5%; Score 233.5; DB 7; Length 331;
Best Local Similarity 42.2%; Pred. No. 2.9e-12;
Matches 49; Conservative 4; Mismatches 38; Indels 25; Gaps 5;

QY 8 PPPHGHHPHGGPPPH-GHPPHGGPPPHPPHGGPPHGH-----PPHG 56
DB 101 PPQGGNQPPPPPPKQGGPPQGGNKPQGGPPPKQGGPPQGGKSPRSPKQPG 160
QY 57 PPPH-GHPPHGGPP-----HGPPHGGPPHGGPPH-GHGHGFHDHGGCDPP 98
DB 161 PPQGGNQPPPPPPKQGGPPQGGNRPQGGPPPKQGGPPQGGKSPRSPQSP 216

RESULT 14

ABO52997
ID ABO52997 standard; protein; 424 AA.

XX AC ABO52997;

DT 09-OCT-2003 (first entry)

DE Human spliceosome associated protein (SAP) #24.

XX Human; SAP; spliceosome associated protein; ribonucleoprotein;
KW RNP complex; RNA affinity substrate; RNP assembly sequence;
KW spliceosomal complex; hnRNP complex; mRNA export complex;
KW mRNA localisation complex; RNA editing complex; intron complex;
KW H complex; telomerase complex; fragile X protein complex;
KW reverse transcriptase complex; gene splicing complex.

OS Homo sapiens.

XX US2003068803-A1.

XX 10-APR-2003.

XX 14-JAN-2002; 2002US-00047991.

XX 12-JAN-2001; 2001US-0261521P.

XX (REED/) REED R.

XX (ZHOU/) ZHOU Z.

PI Reed R, Zhou Z;

XX WPI; 2003-540885/51.

XX Isolating ribonucleoprotein complex, by contacting RNA affinity substrate
PT having ribonucleoprotein assembly sequence and affinity tag, with protein
PT mixture, subjecting complex formed to chromatography, affinity selection.

XX Claim 24; Page; 39pp; English.

XX The invention relates to forming (M1) an isolated ribonucleoprotein (RNP)
CC complex (C), involves contacting an RNA affinity substrate (S) comprising
CC an RNP assembly sequence (AS) and an affinity tag, with a protein mixture

CC to permit formation of (C) on AS, subjecting (C) to chromatographic
CC separation, and subjecting (C) to affinity selection, where the affinity
CC tag (e.g. bacteriophage MS2 coat protein in a fusion protein with E. coli
CC maltose binding protein) binds to an affinity matrix. Also included are
CC an isolated spliceosome preparation (isolated by (M1)), a RNA comprising
CC an RNP complex binding site and at least one phase coat protein (M2) a
CC recognition site, a nucleic acid encoding the RNA, and treating (by
CC subjecting having a disorder associated with abnormal RNP complexes (by
CC obtaining a sample of cells from a subject, purifying RNP complexes from
CC the cells of the subject by (M1), determining the presence in the
CC purified RNP complexes of one or more proteins, and normalising the
CC amount of RNPs in the subject. (M1) is useful for forming an isolated RNP
CC complex selected from a spliceosomal complex (selected from E, A, B and C
CC complex), an hnRNP complex, an mRNA export complex, an mRNA localisation
CC complex, an RNA editing complex, an intron complex, or an H complex. (M1)
CC is useful in a diagnostic assay for determining whether a subject has
CC abnormal RNP complexes, (M2) is useful for treating a subject having a
CC disorder associated with abnormal RNP complexes. (M1) is useful for
CC forming an isolated RNP complex such as a telomerase complex, a fragile X
CC protein complex, a reverse transcriptase complex or a gene splicing
CC complex. The present sequence represents a known human spliceosome
CC associated protein (SAP) isolated by the methods of the invention. Note:
CC The present sequence is not shown in the specification but was obtained
CC from Genbank or Swissprot using the information provided in table 1 of
CC the specification

XX Sequence 424 AA;

Query Match 33.5%; Score 233.5; DB 6; Length 424;
Best Local Similarity 46.0%; Pred. No. 3.5e-12;
Matches 57; Conservative 3; Mismatches 25; Indels 39; Gaps 11;

QY 9 PPHGH-HPHG-PP-----PHG-HHHPHGGPP--HGHPHGGPPHPPHGG 46
DB 286 PGHGHSHPHPPPPGGMPHFGMSQMLAHGHGHPHAGPPGSGGQPPPPGMPHFG 345

QY 47 PPHGHPPHGGPP---PHGHPPHGGPPHGG--HPPHGGPPHGGH-----PPHGHGFHDHGG 94
DB 346 PPMGMPPRGPPPPGSGPMGHP--GMPHGMRGPPPLMPHGYTGPPRPPPYG---YQRG 400

QY 95 CDPP 98

DB 401 LPPP 404

RESULT 15

AAO30246

ID AAO30246 standard; protein; 192 AA.

XX AAO30246;

XX 03-SEP-2003 (first entry)

DE Human Saprip protein fragment #1.

XX Human; forensic analysis; chromosome marker; organelle-specific marker;
KW novel-related disorder; neurological disorder; gene therapy; nootropic;
KW neuroprotective; salivary proline-rich protein; Saprip.

OS Homo sapiens.

OS Synthetic.

XX WO2003046180-A2.

XX 05-JUN-2003.

XX 25-NOV-2002; 2002WO-EP013210.

XX 28-NOV-2001; 2001US-0334147P.

XX 14-DEC-2001; 2001US-0340465P.

XX 18-APR-2002; 2002US-0373947P.

XX (GEST) GENSET SA.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2004, 22:33:30 ; Search time 20.2811 Seconds
(without alignments)
479.034 Million cell updates/sec

Title: US-10-074-225A-6
Perfect score: 697
Sequence: 1 SVNIHRPPPHGHHPHGPPP.....HPPHGHGFHDGFCDPSPSHK 101

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	314	45.1	525	1 KGHUGH	histidine-rich gly
2	260.5	37.4	445	2 A60488	histidine-rich gly
3	259.5	37.2	295	2 B48013	proline-rich prote
4	250	35.9	128	2 D38355	basic proline-rich
5	248	35.6	278	2 B39066	proline-rich prote
6	247.5	35.5	206	1 PIRT3	acidic proline-ric
7	246.5	35.4	227	2 C29149	proline-rich prote
8	246.5	35.4	240	2 A24264	proline-rich prote
9	246.5	35.4	317	2 A28996	proline-rich prote
10	239	34.3	240	2 B24264	proline-rich prote
11	236	33.9	330	2 S22140	nodulin Enod2 - Se
12	235.5	33.8	392	1 PIHUB6	salivary proline-r
13	235.5	33.8	432	2 T06782	extensin - soybean
14	235.5	33.8	580	2 T10863	extensin precursor
15	234.5	33.6	173	2 T51469	glycine/proline-ri
16	233.5	33.5	309	2 S08343	nodulin precursor
17	233.5	33.5	424	2 A54964	spliceosome-associ
18	232.5	33.4	1571	2 T14155	zinc finger protei
19	229.5	32.9	251	1 PIHUPF	salivary proline-r
20	228.5	32.8	188	2 D29149	proline-rich prote
21	225.5	32.4	212	2 B36298	proline-rich prote
22	224	32.1	310	1 PIHUSD	salivary proline-r
23	222.5	31.9	220	2 A36298	proline-rich prote
24	221	31.7	309	2 S10889	proline-rich prote
25	221	31.7	1378	2 T30173	zinc finger protei
26	214.5	30.8	117	2 D40750	proline-rich prote
27	213.5	30.6	301	2 E29149	proline-rich prote
28	210.5	30.2	434	2 S11967	nodule-specific hy
29	210	30.1	297	2 S23737	proline-rich prote

30	210	30.1	440	1 S60755	rhodopsin - Allote
31	210	30.1	620	2 S06733	hydroxyproline-ric
32	209	30.0	97	2 S04554	nodulin - alfalfa
33	208	29.8	223	2 A42817	proline-rich prote
34	207	29.7	727	2 C84534	hypothetical prote
35	206.5	29.6	112	2 S10101	nodulation protein
36	204.5	29.3	539	2 T28770	hypothetical prote
37	202.5	29.1	170	2 A48013	proline-rich prote
38	202	29.0	188	2 JH0481	basic proline-rich
39	201	28.8	452	1 S14332	rhodopsin - northe
40	199.5	28.6	172	2 B29149	proline-rich prote
41	198.5	28.5	1176	2 T49482	hypothetical prote
42	198	28.4	164	2 A30496	proline-rich prote
43	197.5	28.3	260	2 S22373	proline-rich prote
44	197.5	28.3	592	2 T32402	hypothetical prote
45	197.5	28.3	1002	2 T43236	carboxypeptidase C

ALIGNMENTS

RESULT 1

KGHUGH
histidine-rich glycoprotein precursor - human
N;Alternate names: HRG
C;Species: Homo sapiens (man)
C;Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 16-Jun-2000
C;Accession: A01287; S29669
R;Koide, T.; Foster, D.; Yoshitake, S.; Davie, E.W.
Biochemistry 25, 2220-2225, 1986
A;Title: Amino acid sequence of human histidine-rich glycoprotein derived from the nucleolus
A;Reference number: A01287; MUID:86216149; PMID:3011081
A;Accession: A01287
A;Molecule type: mRNA
A;Residues: 1-525 <KOI>
A;Cross-references: GB:AB005803; NID:g2280513; PIDN:BAA21613.1; PID:g2280514
R;Hennis, B.; Havelaar, A.; Klufft, C.
submitted to the EMBL Data Library, October 1991
A;Description: PCR detection of a dinucleotide repeat in the human histidine-rich glycoprotein
A;Reference number: S29669
A;Accession: S29669
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 214-247 <HEN>
A;Cross-references: EMBL:Z17218; NID:g32453; PIDN:CAA78925.1; PID:g32454
C;Comment: Although its physiological function is not yet known, HRG does bind heme, dy
din, and the lysine-binding site of plasminogen. On the basis of its homology with HMW
lood coagulation cascade.
C;Comment: The amino half of this protein is homologous to the first two cystatin-like
ould not have inhibitory activity.
C;Comment: In addition to having a high histidine and proline content, this protein has
e-rich' region.
C;Genetics:
A;Gene: GDB:HRG
A;Cross-references: GDB:120055; OMIM:142640
A;Map position: 3q27-3q27
C;Superfamily: histidine-rich glycoprotein; cystatin homology
C;Keywords: duplication; glycoprotein; heparin binding; tandem repeat
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-525/Product: histidine-rich glycoprotein #status predicted <MAT>
F;19-131/Domain: cystatin homology <CY1>
F;140-246/Domain: cystatin homology <CY2>
F;276-321/Region: proline-rich
F;348-437/Region: histidine-rich
F;351-497/Region: proline-rich
F;63,125,344,345/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;78-89,105-126,218-241/Disulfide bonds: #status predicted

Query Match 45.1%; Score 314; DB 1; Length 525;
Best Local Similarity 41.8%; Pred.No. 1.1e-16;
Matches 64; Conservative 2; Mismatches 23; Indels 64; Gaps 8;

Qy 7 RPP--PHG-----HHPHGPPPHGHHPHGPPP-----HGHP-PHGPPP----- 40

Db 275 KPPFKGSGRDHHPKP-----HEGGPPPPPPDERDHSGLPQQGPPPLPMSCSSCQH 329
 QY 41 -----RHP-----PHGPPPHGHP-----HGPPPHGHPPEGP 67
 Db 330 ATFGTGAQRHSHNNSSDLHPKHSHEQHPPGHGHPHAHPHEHDTHQHPGHHPHGH 389
 QY 68 PPHGHPHPPPPHGHPPHGHGFDHGPCDPPSH 100
 Db 390 HPHGHHPHGHPPHGHPPHGHCHDFQDYGCDPPPH 422

RESULT 2
 A60488
 histidine-rich glycoprotein - bovine (fragments)
 N;Alternate names: autorosette inhibition factor
 C;Species: Bos primigenius taurus (cattle)
 C;Date: 19-Mar-1993 #sequence_revision 23-Mar-1995 #text_change 07-Jul-1995
 C;Accession: S35687; J02196; A60488
 R;Sorensen, C.B.; Krogh-Pedersen, H.; Petersen, T.E.
 FEBS Lett. 328, 285-290, 1993
 A;Title: Determination of the disulphide bridge arrangement of bovine histidine-rich gly
 A;Reference number: S35687; MUID:93351678; PMID:8348977
 A;Accession: S35687
 A;Molecule type: protein
 A;Residues: 1-25;26-52;57-75;82-88;95-119;146-173;175-206;210-309;313-445 <SOR>
 A;Note: 355-Gln and 368-Tyr were also found
 R;Halkier, T.; Andersen, H.; Vestergaard, A.; Magnusson, S.
 Biochem. Biophys. Res. Commun. 200, 78-82, 1994
 A;Title: Bovine histidine-rich glycoprotein is a substrate for bovine plasma factor XIII
 A;Reference number: J02196; MUID:94220160; PMID:7909439
 A;Accession: J02196
 A;Molecule type: protein
 A;Residues: 1-23;35-54,'VK',57-101,'R','TVGEXTEG',116,'N',118,'R',120-136;137-145;150-20

A;Experimental source: plasma
 R;Vestergaard, A.B.; Andersen, H.F.; Magnusson, S.; Halkier, T.
 Thromb. Res. 60, 385-396, 1990
 A;Title: Histidine-rich glycoprotein inhibits contact activation of blood coagulation.
 A;Reference number: A60488; MUID:91196010; PMID:2084959
 A;Accession: A60488
 A;Molecule type: protein
 A;Residues: 1-6,'X',8-15 <VES>
 C;Comment: This protein is a single-chained plasma protein which participates in transgl
 C;Superfamily: histidine-rich glycoprotein; cystatin homology
 C;Keywords: glycoprotein; plasma
 F;2-113/Domain: cystatin homology <CY1>
 F;122-207/Domain: cystatin homology (fragments) <CY2>
 F;7-424,60-71,87-108,165-346,180-203,258-288/Disulfide bonds: #status experimental
 F;74,107,164,266/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 37.4%; Score 260.5; DB 2; Length 445;
 Best Local Similarity 42.4%; Pred. No. 8.7e-13;
 Matches 53; Conservative 4; Mismatches 33; Indels 35; Gaps 6;
 QY 11 HGHHPHGGPP-----HGHPHG-----PPPHGHPHG-----PPP--- 40
 Db 229 HEHSPAGRPFPKPSGSKDHG-HPHESYFRCPPLLEKHNHSDSPFFQARAPLPFPPLGLR 287
 QY 41 -RHPPHGGPPPHGHPH-----GPPPHGHPHPPHPPHGHPPHGHGHPHGHGHPHGHGHPD 96
 Db 288 CPHPFPFGTKGNHRPPHDSHSSDEHNHHPHGHHPHGHHPHGHHPHDPDNDFYDHGPCD 347
 QY 97 PPSHK 101
 Db 348 PPHR 352

RESULT 3
 B48013
 proline-rich proteoglycan 2 precursor, parotid - rat
 C;Species: Rattus norvegicus (Norway rat)
 C;Date: 02-Jun-1995 #sequence_revision 02-Jun-1995 #text_change 20-Aug-1999
 C;Accession: B48013

R;Castle, A.M.; Castle, J.D.
 J. Biol. Chem. 268, 20490-20496, 1993
 A;Title: Novel secretory proline-rich proteoglycans from rat parotid. Cloning and charac
 A;Reference number: A48013; MUID:93388626; PMID:8376404
 A;Accession: B48013
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-295 <CAS>
 A;Cross-references: GB:L17318; NID:g310199; PIDN:AAA03074.1; PID:g310200
 C;Superfamily: proline-rich protein
 C;Keywords: extracellular protein; glycoprotein; tandem repeat

Query Match 37.2%; Score 259.5; DB 2; Length 295;
 Best Local Similarity 38.4%; Pred. No. 7.3e-13;
 Matches 63; Conservative 1; Mismatches 27; Indels 73; Gaps 10;
 QY 2 VNIHRPPPHG---HHPH-----GPPPHG-----HHPHGGPPPHGHP-- 34
 Db 81 VNRPERPPQHGGNHHHPHPPAAGPQRPPQPGSPQGGPPPGGQQRRPPQGGPPPGGQQR 140
 QY 35 -----PHGPPP-----RHPPHGGPPPHGHP-----PHGPPPHGHP--- 71
 Db 141 PPQSGPQGGPPPGGPPQRRPPQGGPPQGGPPQRRPPQGGSPQGGPPPGGPPQRRAPQGGPPPG 200
 QY 72 HP-----PHGP-----PPHGHPPHGHGFDHGPCDPP 98
 Db 201 GPQRPPQSGSPQGGPPPGGPPQRRPPQGGPPPG-----GPQRPP 238

RESULT 4
 D38355
 basic proline-rich peptide IB-8a - human (fragments)
 C;Species: Homo sapiens (man)
 C;Date: 23-Aug-1991 #sequence_revision 23-Aug-1991 #text_change 12-Apr-1995
 C;Accession: D38355
 R;Kaufman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.
 Biochemistry 30, 3351-3356, 1991
 A;Title: Basic proline-rich proteins from human parotid saliva: relationships of the cov
 A;Reference number: A38355; MUID:91190884; PMID:1849422
 A;Accession: D38355
 A;Status: preliminary
 A;Molecule type: protein
 A;Residues: 1-128 <KAU>
 C;Superfamily: proline-rich protein

Query Match 35.9%; Score 250; DB 2; Length 128;
 Best Local Similarity 45.0%; Pred. No. 1.8e-12;
 Matches 50; Conservative 5; Mismatches 36; Indels 20; Gaps 5;
 QY 8 PPHGHHPHGGPPPHGHHPHGGPPH-GHPHGGPPPHHPHGGPPHGH-----PPHGGPPH 60
 Db 10 PPQGGNQPGPPPPPGKPGQPPPPQGGNKPQGGPPPGKPGQPPPPQGGDNKSQSAKPGPPPPQ 69
 QY 61 GHPPHGGPPH-GHPHGGPPPHGH-----PPHGHGHGFDHGPCDPPS 99
 Db 70 GCKPQGGPPPGGNGKPGPPPGKPSARSPPGKPGQPPPGQGGNNPQGP-PPPA 119

RESULT 5
 B39066
 proline-rich protein 15 - rat
 C;Species: Rattus norvegicus (Norway rat)
 C;Date: 20-Mar-1992 #sequence_revision 10-Apr-1992 #text_change 03-May-1996
 C;Accession: B39066
 R;Lin, H.H.; Ann, D.K.
 Genomics 10, 102-113, 1991
 A;Title: Molecular characterization of rat multigene family encoding proline-rich protei
 A;Reference number: A39066; MUID:91257817; PMID:2045095
 A;Accession: B39066
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-278 <LIN>
 A;Cross-references: GB:M36414

QY	7	RPPPHGHHP-----HGPPPHG----HHPHGPPP--HPPHGPPP-----RHPPH	45
D6	33	RPPPSGSQPRPPVNGSQGGPPPPGGPQRPPQGPPPPGGPQRPPPGGPPRPPQ	92
QY	46	GPPPHG---HPPHGPPP--HPPHGPPP--HPPHGPPP--HPPHG---	86
D6	93	GPPPPGGPQPRPPQGPPPPGGPQRPPFQQGPPPPGGPQRPPQGPPPGGPPRPPQGPP	152
QY	87	-HGFHDHGPCDDP	98
D6	153	PGGPQRPPQGPP	165

RESULT 10
B24264
proline-rich protein MP3 - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 09-Sep-1987 #sequence_revision 09-Sep-1987 #text_change 03-May-1996
C:Accession: B24264
R:Ann, D.K.; Carlson, D.M.
J. Biol. Chem. 260, 15863-15872, 1985
A:Title: The structure and organization of a proline-rich protein gene of a mouse multis
A:Reference number: A92508; MUID:86059475; PMID:2999141
A:Accession: B24264
A:Molecule type: DNA
A:Residues: 1-240 <ANN>
C:Superfamily: proline-rich protein

RESULT 11
S22140
modulin Enod2 - Sesbania rostrata
C:Species: Sesbania rostrata
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 20-Aug-1999
C:Accession: S22140
R:de Bruijn, F.J.
submitted to the EMBL Data Library, November 1991
A:Reference number: S22140
A:Accession: S22140
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-330 <DEB>
A:Cross-references: EMBL:X63339; NID:g21372; PIDN:CAA44939.1; PID:g21373
C:Superfamily: proline-rich protein 3

RESULT 12

PIHUB6

salivary proline-rich phosphoprotein precursor PRB1 (large allele) [validated] - human
N;Contains: peptide IB-1; peptide P-E (peptide IB-9); peptide P-F; peptide P-H
C;Species: Homo sapiens (man)
C;Date: 04-Dec-1986 #sequence_revision 12-Apr-1996 #text_change 08-Dec-2000
C;Accession: B40750; A40750; C25372; S02128; S02127; A03293; A90502; A91974; A051974
R;Azen, E.A.; Latreille, P.; Niece, R.L.
Am. J. Hum. Genet. 53, 264-278, 1993
A;Title: PRB1 gene variants coding for length and null polymorphisms among human salivary
A;Reference number: A40750; MUID:93304421; PMID:8317492
A;Accession: B40750
A;Molecule type: DNA
A;Residues: 35-392 <AZE>
A;Cross-references: GB:S62941
A;Experimental source: subject C.J. (large allele)
A;Accession: C40750
A;Molecule type: DNA
A;Residues: 35-127,'R',129-148,'R',150-151,153-187,'K',189-272,'S',274-336,'S',338-392 <AZE>
A;Cross-references: GB:S62929
A;Experimental source: subject M.V.O. (large allele)
A;Accession: A40750
A;Molecule type: DNA
A;Residues: 35-183,245-270,'Q',272-392 <AZ3>
A;Cross-references: GB:S62928
A;Experimental source: subject C.J. (medium allele)
A;Note: authors translated the codon CAA for residue 272 as Arg
R;Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.
J. Biol. Chem. 260, 11123-11130, 1985
A;Title: Differential RNA splicing and post-translational cleavages in the human salivary
A;Reference number: A92492; MUID:85289325; PMID:2993301
A;Accession: C25372
A;Molecule type: mRNA
A;Residues: 1-183,245-392 <MAE>
A;Cross-references: GB:K03204; NID:gl90485; PIDN:AAA60185.1; PID:gl90486
A;Note: alternatively splice forms lacking portions of the repeat region were also found
R;Lyons, K.M.; Stein, J.H.; Smithies, O.
Genetics 120, 267-278, 1988
A;Title: Length polymorphisms in human proline-rich protein genes generated by intragenic
A;Reference number: S02127; MUID:89121440; PMID:2851479
A;Accession: S02128
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 35-127,250-273,'R',275-277,'R',279-336,'S',338-392 <LYO>
A;Cross-references: EMBL:X07517
A;Accession: S02127
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 35-183,245-392 <LY2>
A;Cross-references: EMBL:X07516
R;Kaufman, D.; Hofmann, T.; Bennick, A.; Keller, P.
Biochemistry 25, 2387-2392, 1986
A;Title: Basic proline-rich proteins from human parotid saliva: complete covalent structure
A;Reference number: A90502; MUID:86243355; PMID:3521730
A;Accession: A03293
A;Molecule type: protein
A;Residues: 17-38,'AP',41-51,92-148,'R',150-152 <KA2>
A;Note: among nine basic proline-rich peptides isolated from the saliva, this peptide is
A;Accession: A90502
A;Molecule type: protein
A;Residues: 275-336,'S',338-392 <KAU>
R;Saitoh, E.; Isemura, S.; Sanada, K.
J. Biochem. 94, 1991-1999, 1983
A;Title: Further fractionation of basic proline-rich peptides from human parotid saliva
A;Reference number: A91974; MUID:84161824; PMID:6671974
A;Contents: P-H
A;Accession: A91974
A;Molecule type: protein
A;Residues: 'S',338-392 <SAI>
R;Azen, E.; Lyons, K.M.; McGonigal, T.; Barrett, N.L.; Clements, L.S.; Maeda, N.; Vanin, P.
Proc. Natl. Acad. Sci. U.S.A. 81, 5561-5565, 1984
A;Reference number: A94005; MUID:84298176; PMID:6089212

submitted to the Protein Sequence Database, August 2000

A;Reference number: Z25394

A;Accession: T51469

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-173 <SAT>

A;Residues: I-I/3 <SAI>
A;Cross-references: EMBL:AL391151

A; Cross-references: SMD:AL391151
A; Experimental source: cultivar Columbia; BAC clone K10A8

C;Genetics:
a, experimental source: cultivar Columbia; BAC clone K1UA8

A;Map position: 5

A; Introns: 97/1

A;Note: K10A8 130

Query Match	33.6%;	Score 234.5;	DB 2;	Length 173;
Best Local Similarity	51.7%;	Pred. No. 3.3e-11;		
Matches 45; Conservative	7;	Mismatches 22;	Indels 13;	Gaps 5;
QY	9	PPHGHHPHGPPPHGHHMPHQPPPHGHP-PHGPPPRHRPPhGPPPHGHPHGP-PPHGHPHG 66		
Dd	26	PPHGH-----GYGHHGHC---YGSSYPYPPDP--PPHGYPPVAYPPHGGYPPAGYPPAG 74		

Qy 9 PPHGHHPPPHGHHPPPHGHP-PPGPPPHPPHPPPHGPPPHGPP-PPHGHPPHG 66

Dh 26 PPHG-----GYGHHGHC---YGSSYPYPPPP--PPHGYPPVAYPPHGGYPPAGYPPAG 74

Dbb 26 PPHGH-----GYGHHGHG-----YGSSYPYPPPP--PPHGYPPVAYPPHCGYPPAGYPPAG 74

Qy 67 PPHGHGHPHGPPHGHPPHGHGFHDHG 93
 ||| : | | | | | : | |
Dd 75 YPAGYPAHGYPSHGYPFSPSHSGHHG 101

Db 75 YPPAGYPAHGYP SHGYPRPSHSGHHG 101

Search completed: September 23, 2004, 22:55:57
Job time : 20.2811 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model
Run on: September 23, 2004, 21:27:50 ; Search time 11.7631 Seconds
(without alignments)
447.085 Million cell updates/sec

Title: US-10-074-225A-6
Perfect score: 697
Sequence: 1 SVNIIHRPPPHGHHPGPPP.....HPPHGHGFHDHGPCDPPSHK 101

Scoring table: BLOSUM62
Gapop'10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	697	100.0	526	1	HRG_RABIT
2	314	45.1	525	1	HRG_HUMAN
3	260	37.3	396	1	HRG_BOVIN
4	247.5	35.5	206	1	PRP3_RAT
5	246.5	35.4	261	1	PRP2_MOUSE
6	246.5	35.4	296	1	PMP3_MOUSE
7	233.5	33.5	309	1	NO75_SOYBN
8	233.5	33.5	331	1	PRP1_HUMAN
9	233.5	33.5	424	1	S3B4_HUMAN
10	229.5	32.9	251	1	PRP2_HUMAN
11	228	32.7	247	1	PRB4_HUMAN
12	224	32.1	276	1	PRPL_HUMAN
13	221.5	31.8	234	1	PRPM_HUMAN
14	210.5	30.2	434	1	NO75_LUPLU
15	210	30.1	620	1	EXTN_TOBAC
16	209	30.0	97	1	NO75_MEDSA
17	206.5	29.6	112	1	NO75_PEA
18	205	29.4	815	1	PYGO_DROME
19	201	28.8	439	1	OPSD_LOLSU
20	201	28.8	452	1	OPSD_LOLFO
21	199.5	28.6	172	1	PRP2_RAT
22	197.5	28.3	475	1	S3A2_MOUSE
23	197.5	28.3	1002	1	CBPY_SCHPO
24	192.5	27.6	455	1	OPSD_OCTDO
25	192	27.5	351	1	HRPX_PLALO
26	192	27.5	464	1	S3A2_HUMAN
27	191	27.4	174	1	PRPP_HUMAN
28	190	27.3	166	1	PRPC_HUMAN
29	188.5	27.0	96	1	PRP5_HUMAN
30	187	26.8	215	1	STRD_MOUSE
31	184.5	26.5	373	1	EXT1_ARATH
32	184.5	26.5	464	1	OPSD_SEPOF
33	183	26.3	2716	1	OSA_DROME

34	182.5	26.2	485	1	SSGP_VOLCA
35	180	25.8	1638	1	BRM_DROME
36	179	25.7	1790	1	SEPA_EMENI
37	176.5	25.3	1097	1	S24C_ARATH
38	175	25.1	306	1	EXTN_DAUCA
39	175	25.1	639	1	SF01_HUMAN
40	175	25.1	653	1	SF01_MOUSE
41	174	25.0	346	1	PRF1_LYCES
42	173	24.8	309	1	HXA4_CHICK
43	173	24.8	462	1	ANX7_DICDI
44	172.5	24.7	431	1	ACRO_RABIT
45	172	24.7	134	1	PRL5_HUMAN

ALIGNMENTS

RESULT 1

HRG_RABIT	STANDARD;	PRT;	526 AA.
AC Q28640;			
DT 01-NOV-1997 (Rel. 35, Created)			
DT 01-NOV-1997 (Rel. 35, Last sequence update)			
DT 15-MAR-2004 (Rel. 43, Last annotation update)			
DE Histidine-rich glycoprotein precursor (Histidine-proline rich glycoprotein) (HPRG) (Fragment).			
GN HRG.			
OS Oryctolagus cuniculus (Rabbit).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.			
OX NCBI_TaxID=9986;			
RN [1]			
RP SEQUENCE FROM N.A., AND SEQUENCE OF 9-23; 301-313 AND 422-429.			
RC TISSUE=Serum;			
RX MEDLINE=96229917; PubMed=8639676;			
RA Borza D.-B., Tatum F.M., Morgan W.T.;			
RT "Domain structure and conformation of histidine-proline-rich glycoprotein.";			
RL Biochemistry 35:1925-1934(1996).			
CC -!- FUNCTION: The physiological function is not yet known. It binds heme, dyes and divalent metal ions. It can inhibit rosette formation and is known to interact with heparin, thrombospondin, and the lysine-binding site of plasminogen. On the basis of its homology with HMW kininogen, the His-rich region of this protein may mediate the contact activation phase of intrinsic blood coagulation cascade.			
CC -!- SUBCELLULAR LOCATION: Secreted.			
CC -!- TISSUE SPECIFICITY: Expressed by the liver and secreted in plasma.			
CC -!- DOMAIN: In addition to having a high His and Pro content, this protein has many internal repeats. 15 tandem repetitions of a 5-residue sequence (G[H/P][H/P]PH, consensus) form a His/Pro-rich region.			
CC -!- SIMILARITY: Contains 2 cystatin-like domains.			
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC -----			
CC EMBL; U32189; AAC48516.1; -			
CC InterPro; IPR000010; Cystatin.			
CC Pfam; PF00031; cystatin; 1.			
CC SMART; SM00043; CY; 2.			
CC Glycoprotein; Heparin-binding; Repeat; Signal.			
CC NON TER 1 8			
CC SIGNAL <1 9 526			
CC CHAIN 9 126			
CC DOMAIN 127 243			
CC DOMAIN 251 296			
CC POTENTIAL. HISTIDINE-RICH GLYCOPROTEIN.			
CC CYSTATIN-LIKE 1.			
CC CYSTATIN-LIKE 2.			
CC PRO-RICH.			


```
ID PRP2_MOUSE STANDARD; PRT; 261 AA.
AC P05142;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Proline-rich protein MP-2 precursor.
GN PRH1 OR PRP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86059475; PubMed=2999141;
RA Ann D.K., Carlson D.M.;
RT "The structure and organization of a proline-rich protein gene of a
mouse multigene family."
RL J. Biol. Chem. 260:15863-15872 (1985).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M12099; AAA40004.1; -.
DR MGD; MGI:97773; Prhl.
KW Repeat; Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 261 PROLINE-RICH PROTEIN MP-2.
SQ SEQUENCE 261 AA; 26034 MW; 36E13BA7387F47D4 CRC64;

Query Match 35.4%; Score 246.5; DB 1; Length 261;
Best Local Similarity 43.6%; Pred. No. 7.7e-11;
Matches 58; Conservative 0; Mismatches 34; Indels 41; Gaps 9;

QY 7 RPPPHGHHP-----HGPPPHG-----HHPHGPPPHG-----HPPHGPPP 45
Db 33 RPPSGFQPRPPVNGSQGGPPPPGPPQPRPPGPPPPGPPPPGPPPPGPPPPQ 92
QY 46 GPPPHG-----HPPHGPPPHG-----HPPHGPPPHG-----HPPHG--- 86
Db 93 GPPPGGQPRPPQPPGPPGPPGPPQPRPPGPPGPPQPRPPGPPGPPGLRPPQGGPPP 152
QY 87 -HGFDHGHGCDPP 98
Db 153 PAGPQPRPPQGGPP 165

RESULT 6
PMP3_MOUSE STANDARD; PRT; 296 AA.
ID PMP3_MOUSE
AC P05143;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Proline-rich protein MP-3 (Fragment).
GN PRH1 OR PRP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86059475; PubMed=2999141;
RA Ann D.K., Carlson D.M.;
RT "The structure and organization of a proline-rich protein gene of a
mouse multigene family."
RL J. Biol. Chem. 260:15863-15872 (1985).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M12100; AAA40005.1; -.
DR MGD; MGI:97773; Prhl.
KW Repeat.
FT NON TER 1 1
SQ SEQUENCE 296 AA; 29521 MW; 7F146824E8AF3269 CRC64;

Query Match 35.4%; Score 246.5; DB 1; Length 296;
Best Local Similarity 43.6%; Pred. No. 8.5e-11;
Matches 58; Conservative 0; Mismatches 34; Indels 41; Gaps 9;

QY 7 RPPPHGHHP-----HGPPPHG-----HHPHGPPPHG-----HPPHGPPP 45
Db 12 RPPSGSQPRPPVNGSQGGPPPPGPPQPRPPGPPPPGPPPPGPPPPGPPPPQ 71
QY 46 GPPPHG-----HPPHGPPPHG-----HPPHGPPPHG-----HPPHG--- 86
Db 72 GPPPGGQPRPPQPPGPPGPPGPPQPRPPGPPGPPGPPQPRPPGPPGPPQPRPPQGGPPP 131
QY 87 -HGFDHGHGCDPP 98
Db 132 PGGPQPRPPQGGPP 144

RESULT 7
NO75_SOYBN STANDARD; PRT; 309 AA.
ID NO75_SOYBN
AC P08297;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Early nodulin 75 precursor (N-75) (NGM-75).
GN ENOD2A AND ENOD2B.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Wayne;
RX MEDLINE=91322483; PubMed=2101308;
RA Franssen H.J., Thompson D.V., Idler K., Kormelink R., van Kammen A.,
RA Bisseling T.;
RT "Nucleotide sequence of two soybean ENOD2 early nodulin genes
encoding Ngm-75."
RL Plant Mol. Biol. 14:103-106 (1990).
RN [2]
RP SEQUENCE OF 69-309 FROM N.A.
RA Franssen H.J., Nap J.-P., Gloudemans T., Stiekema W., van Dam H.,
RA Govers F., Louwerse J., van Kammen A., Bisseling T.;
RT "Characterization of cDNA for nodulin-75 of soybean: a gene product
involved in early stages of root nodule development."
RL Proc. Natl. Acad. Sci. U.S.A. 84:4495-4499 (1987).
CC -!- FUNCTION: NODULIN N-75 IS INVOLVED IN EARLY STAGES OF ROOT NODULE
DEVELOPMENT.
CC -!- INDUCTION: During nodulation in legume roots after Rhizobium
infection.
CC -!- SIMILARITY: TO OTHER PLANTS N-75.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```


RN [3]
RP CHARACTERIZATION OF THE SPLICEOSOME.
RX MEDLINE=20337962; PubMed=10882114;
RA Das R., Zhou Z., Reed R.;
RT "Functional association of U2 snRNP with the ATP-independent
RT spliceosomal complex E";
RL Mol. Cell 5:779-787(2000).
CC -!- FUNCTION: Subunit of the splicing factor SF3B required for 'A'
CC complex assembly formed by the stable binding of U2 snRNP to the
CC branchpoint sequence (BPS) in pre-mRNA. Sequence independent
CC binding of SF3A/SF3B complex upstream of the branch site is
CC essential, it may anchor U2 snRNP to the pre-mRNA. May also be
CC involved in the assembly of the 'E' complex. SF3B4 has been
CC found in complex 'B' and 'C' as well. Belongs also to the
CC minor U12-dependent spliceosome, which is involved in the splicing
CC of rare class of nuclear pre-mRNA intron.
CC -!- SUBUNIT: Component of splicing factor SF3B which is composed of
CC four subunits; SF3B4/SAP49, SF3B3/SAP130, SF3B2/SAP145,
CC SF3B1/SAP155. SF3B associates with the splicing factor SF3A and a
CC 12S RNA unit to form the U2 small nuclear ribonucleoproteins
CC complex (U2 snRNP). SF3B4 interacts directly with SF3B2.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: STRONG, TO C.ELEGANS C08B11.5 AND, PARTIAL, TO
CC YEAST HSH49.
CC -!- SIMILARITY: Contains 2 RNA recognition motif (RRM) domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L35013; AAA60300.1; -;
DR EMBL; BC004273; AAH04273.1; -;
DR EMBL; BC013886; AAH13886.1; -;
DR PIR; A54964; A54964.
DR HSSP; P11940; 1CVJ.
DR Genew; HGNC:10771; SF3B4.
DR GK; Q15427; -;
DR MIM; 605593; -;
DR GO; GO:0005681; C:spliceosome complex; TAS.
DR GO; GO:0008248; F:pre-mRNA splicing factor activity; TAS.
DR GO; GO:0006397; P:mRNA processing; TAS.
DR GO; GO:0006371; P:mRNA splicing; TAS.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS0102; RRM; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Spliceosome; mRNA processing; mRNA splicing; Nuclear protein;
KW RNA-binding; Repeat.
FT DOMAIN 13 91 RNA-BINDING (RRM) 1.
FT DOMAIN 100 179 RNA-BINDING (RRM) 2.
FT DOMAIN 215 218 POLY-PRO.
FT DOMAIN 262 268 POLY-PRO.
FT SEQUENCE 424 AA; 44385 MW; 212472A25D3FF002 CRC64;
SQ
Query Match 33.5%; Score 233.5; DB 1; Length 424;
Best Local Similarity 46.0%; Pred. No. 8.5e-10;
Matches 57; Conservative 3; Mismatches 25; Indels 39; Gaps 11;
QY 9 PPHGH-HPHG-PP-----PHG-HHPHGPPP--HGHPHGPPRHPPHG 46
DB 286 PGGHSHPHPPPPGGMPPHPCMSQMLAHGHGPHGLGHAGPGSGGQPPRPPGMPHPG 345
QY 47 PPHGHHPHGPP-PPHGHPHGP-PPHG--HPPHGPPPHGH-----PPHGHGFDHGP 94
DB 346 PPMGMPRGPFGSPMGHP--GPMPPHGMGRGPPPLMPPHGYTGPPRPPPYG---YQGRP 400
QY 95 CDDP 98
||

Db 401 LPPP 404
RESULT 10
PRP2_HUMAN
ID PRP2_HUMAN STANDARD; PRT; 251 AA.
AC P02812;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Salivary proline-rich protein precursor (Clone CP7) [Contains: Basic
DE peptide P-F] (Fragment).
GN PRB2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=85289325; PubMed=2993301;
RA Maeda N., Kim H.-S., Azen E.A., Smithies O.;
RT "Differential RNA splicing and post-translational cleavages in the
RT human salivary proline-rich protein gene system.";
RL J. Biol. Chem. 260:11123-11130(1985).
RN [2]
SEQUENCE OF 134-194.
RX MEDLINE=83265674; PubMed=6874669;
RA Saitoh E., Isemura S., Sanada K.;
RT "Complete amino acid sequence of a basic proline-rich peptide, P-F,
RT from human parotid saliva";
RL J. Biochem. 93:883-888(1983).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K03208; AAA60189.1; -;
DR PIR; E25372; PIHUPF.
DR Genew; HGNC:9338; PRB2.
DR MIM; 168810; -;
KW Repeat; Parotid gland; Multigene family.
FT NON_TER 1 1
FT CHAIN 134 194 BASIC PEPTIDE P-F.
FT REPEAT <1 9
FT REPEAT 10 71
FT REPEAT 72 133
FT REPEAT 134 195
FT SEQUENCE 251 AA; 24641 MW; D779F590C0EBF30B CRC64;
SQ
Query Match 32.9%; Score 229.5; DB 1; Length 251;
Best Local Similarity 43.6%; Pred. No. 1e-09;
Matches 48; Conservative 3; Mismatches 34; Indels 25; Gaps 5;
QY 8 PPHGHHPHGPPPHGHHPHGPPPH-GHPHGHPPRHPHGHPPPHGH-----PPH 55
DB 19 PPQGNQPPGPPPPPGKPPGPPPPQGNKPPGPPPGKPPGPPPPQGNKSQSARSPPGKPQ 78
QY 56 GPPPH-GHPHGHPPPHGHHPHGPPPHGH-----PPHGHGFDHGHGCDPP 98
DB 79 GPPPGGNQPPGPPPPPGKPPGPPPPQGNKPPGPPPGKPPGPPPPGK-----PQGPP 122
RESULT 11
PRB4_HUMAN
ID PRB4_HUMAN STANDARD; PRT; 247 AA.
AC P10163; P02813;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)


```
RT proline-rich proteins to specific loci.";
RL Genetics 120:255-265(1988).
RN [2]
RP SEQUENCE OF 165-234.
RX MEDLINE=83186122; PubMed=6841349;
RA Saitoh E., Isemura S., Sanada K.;
RT "Complete amino acid sequence of a basic proline-rich peptide, P-D,
RT from human parotid saliva.";
RL J. Biochem. 93:495-502(1983).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X07704; CAA30542.1; -.
DR MIM; 168730; -.
DR MIM; 180990; -.
DR GO; GO:0005576; C:extracellular; NAS.
KW Repeat; Parotid gland; Multigene family.
FT NON_TER 1
FT CHAIN 165 234 PEPTIDE P-D.
SQ SEQUENCE 234 AA; 23676 MW; 310AFF13A4E747F CRC64;
Query Match 31.8%; Score 221.5; DB 1; Length 234;
Best Local Similarity 42.3%; Pred. No. 3.4e-09;
Matches 47; Conservative 5; Mismatches 40; Indels 19; Gaps 5;
QY 7 RPPPHGHHPGPPH-GHHPGPPPHGHPGPPHPR--HPPHGGPPH-----GH 52
Db 15 RPPPPGKPGPPPGGNGSQGPPPHGKPERPPPGGNGSQGPPHKGKPERPPPGGN 74
QY 53 PHGPPPHGHPGPPH-GHPPHGGPPH-----GHPPHGHGHDHGGCDPP 98
Db 75 QSQGPPPTGKPEGPPPGGNGSQGPPPHGKPERPPPGGNGSQGSHRPPPPP 125
RESULT 14
NO75_LUPLU STANDARD; PRT; 434 AA.
AC Q06841;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Early nodulin 75 protein (N-75) (NGM-75) (Fragment).
GN ENOD2.
OS Lupinus luteus (Yellow lupine).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Genisteae; Lupinus.
OX NCBI_TaxID=3873;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Root nodules;
RX MEDLINE=91355883; PubMed=2103455;
RA Szczyglowski K., Legocki A.B.;
RT "Isolation and nucleotide sequence of cDNA clone encoding nodule-
RT specific (hydroxy)proline-rich protein LENOD2 from yellow lupin.";
RL Plant Mol. Biol. 15:361-363(1990).
CC -!- FUNCTION: INVOLVED IN EARLY STAGES OF ROOT NODULE DEVELOPMENT.
CC -!- SIMILARITY: TO OTHER PLANTS N-75.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
DR EMBL; X55371; CAA39050.1; -.
DR FIR; S11967; S11967.
KW Nodulation; Repeat.
FT NON_TER 1
FT DOMAIN 6 407 36 X 8 AA APPROXIMATE REPEATS OF
FT P-P-P-H-E-K-P-P.
SQ SEQUENCE 434 AA; 49827 MW; 8CE8E140C21B6664 CRC64;
Query Match 30.2%; Score 210.5; DB 1; Length 434;
Best Local Similarity 37.0%; Pred. No. 3.1e-08;
Matches 51; Conservative 6; Mismatches 40; Indels 41; Gaps 8;
QY 4 IIRPPPHG-----HHPHG-----PPPHGHP-HGPPPHGHPGPPPHR-----P 43
Db 137 LVH-PPPHDKPPIEYHPHKEKPPPIEYPPHKEKPPPIEYPPHKEKPPVYEP 195
QY 44 PHGPPPHGHP-----HGPPPHGHPGPPPHGHP-HGPPPHGHP 84
Db 196 PYEKPPPVHPPPPDEKPPPIEYPPPLEKPPVHEPPYEKPPPAQPPPHDKPPIEYPPHKEKPP 255
QY 85 -HGHGFHDHGCDPPSHK 101
Db 256 VYEPPIERSPPVHPPSHE 273
RESULT 15
EXTN TOBAC STANDARD; PRT; 620 AA.
ID -EXTN TOBAC
AC P13983;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Extensin precursor (Cell wall hydroxyproline-rich glycoprotein).
GN HRPNT3.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC lamids; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Xanthi; TISSUE=Leaf;
RX MEDLINE=90128263; PubMed=2612909;
RA Keller B., Lamb C.J.;
RT "Specific expression of a novel cell wall hydroxyproline-rich
RT glycoprotein gene in lateral root initiation.";
RL Genes Dev. 3:1639-1646(1989).
CC -!- FUNCTION: Has a specialized structural function, possibly in
CC the mechanical penetration of the cortex and epidermis of the
CC main root.
CC -!- SUBCELLULAR LOCATION: Extracellular matrix.
CC -!- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
CC GLYCOSYLATED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X13885; CAA32090.1; -.
DR FIR; S06733; S06733.
DR InterPro; IPR000480; Glutelin.
DR PRINTS; PR00211; GLUTELIN.
KW Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
KW Hydroxylation.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 620 EXTENSIN.
FT REPEAT 70 73 H-A-P-P.
FT REPEAT 148 151 H-A-P-P.
```

FT DOMAIN 229 242 2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.
FT REPEAT 229 235 1.
FT REPEAT 236 242 2.
FT DOMAIN 205 620 CONTAINS THE SER-PRO(4) REPEATS.
FT DOMAIN 499 600 3 X APPROXIMATE TANDEM REPEATS.
SQ SEQUENCE 620 AA; 65406 MW; 641DD2278AB28524 CRC64;

Query Match 30.1%; Score 210; DB 1; Length 620;
Best Local Similarity 42.3%; Pred. No. 4.4e-08;
Matches 55; Conservative 4; Mismatches 29; Indels 42; Gaps 13;

Qy 8 PPHGH-----HPHGGPPPHGHHP---HGPPPH-GH-PP---HGPPPR---HPPHG 46
Db 59 PPSRGHVPSRPHAPPRHAYPPPSHGLPPSVGGPPPHRGHLPPSRGFENPPSPVISPSHP 118

Qy 47 PPHGHPP--HGP---PPHGHPP-----HGPPPHGH-PPHGGPPPHGH---PPHGHGFH 90
Db 119 PPSYGAPPPSHGPHLPSHGQRPSPSPSHGHAPPSGGHTPPRGQHPPSHRRRPPSPSRHG-- 176

Qy 91 DHGPCDPPSH 100
Db 177 -HPP--PTY 183

Search completed: September 23, 2004, 22:50:36
Job time : 12.7631 secs

Blank

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

QM protein - protein search, using sw model

Run on: September 23, 2004, 22:25:45 ; Search time 59.6265 Seconds
(without alignments)
534.449 Million cell updates/sec

Title: US-10-074-225A-6
Perfect score: 697
Sequence: 1 SVNIIHRPPPHGHHPHGGPP.....HPPHGHGFHDHGCDPPPSHK 101

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	420.5	60.3	546	6 Q9BGU1	Q9bgul bos taurus
2	375	53.8	525	11 Q99PS8	Q99ps8 rattus norv
3	333	47.8	525	11 Q99PS6	Q99ps6 mus musculu
4	333	47.8	525	11 Q9ESB3	Q9esb3 mus musculu
5	333	47.8	525	11 Q99PS5	Q99ps5 mus musculu
6	323	46.3	515	11 Q99PS7	Q99ps7 rattus norv
7	315	45.2	510	11 Q9ESB2	Q9esb2 rattus norv
8	259.5	37.2	295	11 Q07611	Q07611 rattus norv
9	258.5	37.1	511	6 Q95JD0	Q95jd0 sus scrofa
10	258.5	37.1	566	6 Q95JD1	Q95jd1 sus scrofa
11	254.5	36.5	676	6 Q95JC9	Q95jc9 sus scrofa
12	246.5	35.4	227	11 Q62107	Q62107 mus musculu
13	246.5	35.4	317	11 Q62103	Q62103 mus musculu
14	246	35.3	274	11 Q04154	Q04154 rattus norv
15	242	34.7	112	11 Q9CR74	Q9cr74 mus musculu
16	236	33.9	330	10 Q41402	Q41402 sesbania ro

17	235.5	33.8	358	4 Q08805	Q08805 homo sapien
18	235.5	33.8	432	10 Q39835	Q39835 glycine max
19	235.5	33.8	580	10 Q09083	Q09083 phaseolus v
20	234.5	33.6	173	10 Q9LF59	Q9lf59 arabidopsis
21	233.5	33.5	297	4 Q16038	Q16038 homo sapien
22	233.5	33.5	338	4 Q86YA1	Q86ya1 homo sapien
23	233.5	33.5	424	11 Q8QZY9	Q8qzy9 mus musculu
24	232.5	33.4	1571	11 Q54978	Q54978 mus musculu
25	232.5	33.4	1572	11 Q8R5N0	Q8r5n0 mus musculu
26	232.5	33.4	1617	11 Q8OU47	Q8ou47 mus musculu
27	231	33.1	173	4 Q15214	Q15214 homo sapien
28	230.5	33.1	382	4 Q00599	Q00599 homo sapien
29	228.5	32.8	102	5 Q9VUE1	Q9vuel drosophila
30	228.5	32.8	188	11 Q62106	Q62106 mus musculu
31	225	32.3	80	5 Q9VPE1	Q9vpz1 drosophila
32	224	32.1	238	4 Q00600	Q00600 homo sapien
33	223	32.0	594	5 Q9VEP4	Q9vep4 drosophila
34	222.5	31.9	244	10 Q9FUR6	Q9fur6 cladastis
35	221	31.7	309	4 Q04118	Q04118 homo sapien
36	221	31.7	1378	11 Q61138	Q61138 mus musculu
37	218.5	31.3	269	10 Q9FUR7	Q9fur7 styphnolobi
38	213.5	30.6	301	11 Q62105	Q62105 mus musculu
39	212.5	30.5	504	11 Q91X93	Q91x93 mus musculu
40	212.5	30.5	1749	5 Q86IN4	Q86in4 dictyosteli
41	211.5	30.3	956	10 Q9LJ64	Q9lj64 arabidopsis
42	211	30.3	233	5 Q8MP30	Q8mp30 dictyosteli
43	211	30.3	472	3 Q59907	Q59907 neurospora
44	210	30.1	297	10 Q41122	Q41122 phaseolus v
45	209.5	30.1	205	10 Q9FI79	Q9fi79 arabidopsis

ALIGNMENTS

RESULT 1

Q9BGU1 ID Q9BGU1 PRELIMINARY; PRT; 546 AA.
AC Q9BGU1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Histidine-rich glycoprotein.
GN BTHRG.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,
RA Ohashi T., Sato N., Hirata D., Tsuchida N., Koide T.;
RT "Molecular diversity of mammalian histidine-rich glycoprotein.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AB055894; BAB33091.1; --
DR GO; GO:0004869; F:cysteine protease inhibitor activity; IEA.
DR InterPro; IPR000010; Cystatin.
DR Pfam; PF00031; cystatin; 1.
DR SMART; SM00043; CY; 2.
SQ SEQUENCE 546 AA; 61948 MW; 26264858824D89EE CRC64;

Query Match 60.3%; Score 420.5; DB 6; Length 546;
Best Local Similarity 63.8%; Pred. No. 1.1e-32;
Matches 67; Conservative 4; Mismatches 21; Indels 13; Gaps 2;

Qy	6	HRPP-----PHGHHPGPPPHGHHPGPPPHGHHPGPPPHGHHPGPPPHGHHPG 56
Dd	341	HRPPHDHSSDEHHFHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGH 396
Qy	57	PPPHGHPPPHGPPPHGPPPHGPPPHGPPPHGPPPHGPPPHGPPPHGPPPHGPPSHK 101
Dd	397	HHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHHR 441

Matches 58; Conservative 0; Mismatches 34; Indels 41; Gaps 9;	
QY	7 RPPPHGHHP-----HGPPPHG-----HHPHGPPPHG-----HPPHGPPP-----RHPPH 45
Db	33 RPPPSGFQPRPFVNGSQGGPPPPGGPQPRPPQGGPPPPGGPQPRPPQGGPQPRPPQ 92
QY	46 GPPPHG-----HPPHGPPPHG-----HPPHGPPPHG-----HPPHGPPPHG-----HPPHG--- 86
Db	93 GPPPPGGPQPRPPQGGPPPPGGPQPRPPQGGPPPPGGPQPRPPQGGPQPRPPQGGPQPRPPQ 152
QY	87 -HGFDHGGPCDPP 98
Db	153 PAGPQPRPPQGGP 165
RESULT 13	
ID	Q62103 PRELIMINARY; PRT; 317 AA.
AC	Q62103;
DT	01-NOV-1996 (TrEMBLrel. 01, Created)
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	Proline-rich protein precursor.
GN	PRP2 OR PRP.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_TaxID=10090;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CD-1; TISSUE=Liver;
RX	MEDLINE=88273214; PubMed=2839509;
RA	Ann D.K., Smith M.K., Carlson D.M.;
RT	"Molecular evolution of the mouse proline-rich protein multigene family. Insertion of a long interspersed repeated DNA element.";
RL	J. Biol. Chem. 263:10887-10893(1988).
DR	EMBL; M23236; AAA53048.1; --
DR	PIR; A28996; A28996.
DR	MGD; MGI:1932491; Prp2.
KW	Signal.
FT	SIGNAL. 1 15 POTENTIAL.
FT	CHAIN 16 317 PROLINE-RICH PROTEIN.
SQ	SEQUENCE 317 AA; 31719 MW; 019301BE31D73278 CRC64;
Query Match 35.4%; Score 246.5; DB 11; Length 317;	
Best Local Similarity 43.6%; Pred. No. 2.8e-16;	
Matches 58; Conservative 0; Mismatches 34; Indels 41; Gaps 9;	
QY	7 RPPPHGHHP-----HGPPPHG-----HHPHGPPPHG-----HPPHGPPP-----RHPPH 45
Db	33 RPPPSGSQPRPFVNGSQGGPPPPGGPQPRPPQGGPPPPGGPQPRPPQGGPQPRPPQ 92
QY	46 GPPPHG-----HPPHGPPPHG-----HPPHGPPPHG-----HPPHGPPPHG-----HPPHG--- 86
Db	93 GPPPPGGPQPRPPQGGPPPPGGPQPRPPQGGPPPPGGPQPRPPQGGPQPRPPQGGPQPRPPQ 152
QY	87 -HGFDHGGPCDPP 98
Db	153 PGGPQPRPPQGGP 165
RESULT 14	
ID	Q04154 PRELIMINARY; PRT; 274 AA.
AC	Q04154;
DT	01-NOV-1996 (TrEMBLrel. 01, Created)
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	Salivary proline-rich protein.
GN	RP15.
OS	Rattus norvegicus (Rat).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
Query Match 34.7%; Score 242; DB 11; Length 112;	
Matches 57; Conservative 1; Mismatches 29; Indels 34; Gaps 9;	
QY	8 PPHHG--HHPHGPPPHG--HHPHGPPPHGHP---PHGPPP-----RHPPHGPPPHGHP- 53
Db	104 PPPGGPQPRPPQGGPPPPGGPQGGPQGGPQGGPQGGPQGGPQGGPQGGPQGGPQ 163
QY	54 --PHGPPPHG--HPPHGPPPHGHP---PHGPPPHGHPHGHGHDHP-----CDP 97
Db	164 QGPQGGPPPPGGPQGGPQGGPQGGPQGGPQGGPQGGPQGGPQGGPQGGPQGGP 219
QY	98 P 98
Db	220 P 220
RESULT 15	
Q9CR74	PRELIMINARY; PRT; 112 AA.
ID	Q9CR74
AC	Q9CR74;
DT	01-JUN-2001 (TrEMBLrel. 17, Created)
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE	EST AA238765.
GN	1810007E14RIK OR AA238765.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_TaxID=10090;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=C57BL/6J; TISSUE=Pancreas, and Small intestine;
RX	MEDLINE=21085660; PubMed=11217851;
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA	Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA	Hayashizaki Y.;
RT	"Functional annotation of a full-length mouse cDNA collection.";
RL	Nature 409:685-690(2001).
DR	EMBL; AK013007; BAB31507.1; --
DR	EMBL; AK008311; BAB25594.1; --
DR	MGD; MGI:1858382; 1810007E14RIK.
SQ	SEQUENCE 112 AA; 10637 MW; 16D773538A95A230 CRC64;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2004, 22:47:25 ; Search time 18.253 Seconds
(without alignments)
285.664 Million cell updates/sec

Title: US-10-074-225A-6
Perfect score: 697
Sequence: 1 SVNIIHRPPPHGHHPHPPPP.....HPPHGHGFHDHGPCDPPPSHK 101

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	314	45.1	525	4	US-09-976-594-64
2	229.5	32.9	124	2	US-08-925-237-2
3	215.5	30.9	104	4	US-09-547-693-235
4	208	29.8	865	4	US-09-281-766-19
5	196	28.1	82	2	US-08-925-237-4
6	175	25.1	306	1	US-08-217-327-6
7	171	24.5	214	1	US-08-217-327-4
8	167	24.0	1958	1	US-07-945-283-2
9	164	23.5	167	3	US-09-507-323B-13
10	161	23.1	79	1	US-08-487-359-3
11	161	23.1	79	1	US-08-222-798A-3
12	160.5	23.0	78	1	US-08-487-359-5
13	160.5	23.0	78	1	US-08-222-798A-5
14	160.5	23.0	144	1	US-08-642-255-49
15	160.5	23.0	720	3	US-09-219-849-4
16	160.5	23.0	777	1	US-08-642-255-53
17	160	23.0	79	1	US-08-487-359-4
18	160	23.0	79	1	US-08-222-798A-4
19	159	22.8	79	1	US-08-487-359-1
20	159	22.8	79	1	US-08-487-359-8
21	159	22.8	79	1	US-08-222-798A-1
22	159	22.8	79	1	US-08-222-798A-8
23	158.5	22.7	148	4	US-09-461-325-453
24	158.5	22.7	148	4	US-10-012-542-453
25	158.5	22.7	234	1	US-08-642-255-51
26	158.5	22.7	504	3	US-09-219-849-3
27	158.5	22.7	561	1	US-08-642-255-52

28 158 22.7 79 1 US-08-487-359-2 Sequence 2, Appli
29 158 22.7 79 1 US-08-487-359-6 Sequence 6, Appli
30 158 22.7 79 1 US-08-222-798A-2 Sequence 2, Appli
31 158 22.7 79 1 US-08-222-798A-6 Sequence 6, Appli
32 157 22.5 76 4 US-09-547-693-233 Sequence 233, App
33 155 22.2 590 1 US-08-021-608D-8 Sequence 8, Appli
34 155 22.2 590 1 US-08-726-160-8 Sequence 8, Appli
35 155 22.2 590 5 PCT-US94-01782-8 Sequence 8, Appli
36 155 22.2 643 1 US-08-021-608D-10 Sequence 10, Appli
37 155 22.2 643 1 US-08-726-160-10 Sequence 10, Appli
38 155 22.2 643 5 PCT-US94-01782-10 Sequence 10, Appli
39 155 22.2 644 1 US-08-021-608D-2 Sequence 2, Appli
40 155 22.2 644 1 US-08-726-160-2 Sequence 2, Appli
41 155 22.2 644 5 PCT-US94-01782-2 Sequence 2, Appli
42 151.5 21.7 330 1 US-08-642-255-32 Sequence 32, Appli
43 151.5 21.7 408 1 US-07-609-716-65 Sequence 65, Appli
44 151.5 21.7 408 3 US-08-475-411A-65 Sequence 65, Appli
45 151.5 21.7 408 3 US-08-478-029A-65 Sequence 65, Appli

ALIGNMENTS

RESULT 1
US-09-976-594-64
; Sequence 64, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 64
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 085596CD1
US-09-976-594-64

Query Match 45.1%; Score 314; DB 4; Length 525;
Best Local Similarity 41.8%; Pred. No. 3.7e-21;
Matches 64; Conservative 2; Mismatches 23; Indels 64; Gaps 8;

Qy 7 RPP--PHG-----HHPHGPPPHGHHPHGP-----HGHP-PHGP-----40

Db 275 KPPFKPHGSRDHHHPKP-----HEHGPPPPPPDERDHSGLPQQPPPLPMSCSCQH 329

Qy 41 -----RHP-----PHGPPPHGHP-----HGPPPHGHP 67

Db 330 ATFGTNGAQRSSNNSSDLPHKHSHEQHHPGHHPHAHPHEHDTHRQHPHHPHGH 389

Qy 68 PPHGHPHPPHPPHPPHPPHPPHPPHPPHPPHPPHPPHPPHPPHPPHPPHPPH 100

Db 390 HPHGHPHPPHPPHPPHPPHPPHPPHPPHPPHPPHPPHPPHPPHPPHPPH 422

RESULT 2

US-08-925-237-2
; Sequence 2, Application US/08925237
; Patent No. 5981720
; GENERAL INFORMATION:
; APPLICANT: Azen, Edwin A.
; APPLICANT: Pan, David
; TITLE OF INVENTION: Human Salivary Proteins And Fragments
; TITLE OF INVENTION: Thereof Having Alpha-Glucosidase Inhibitory Activity


```

;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 257-3501
; TELEFAX: (608) 283-2275
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 82 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-925-237-4
;
; Query Match 28.1%; Score 196; DB 2; Length 82;
; Best Local Similarity 52.9%; Pred. No. 2.6e-11;
; Matches 37; Conservative 4; Mismatches 25; Indels 4; Gaps 3;
;
; QY 15 PHGPPPH-GHHPHGPPPHGHPPHGGPPR--HPPHGGPPPHGHPPHGGPPPHGHPPHGGPPPHG 71
; Db 6 PQGPPPPQGGNPPQGGPPPPPGKPPQGGNPPQGGNPPQGGNPPQGGNPPQGGNPPQGGNPPG 65
;
; QY 72 HPPHGGPPPHG 81
; Db 66 -KPQGGPPPG 74
;
;
; RESULT 6
; US-08-217-327-6
; Sequence 6, Application US/08217327
; Patent No. 5474925
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E
; APPLICANT: Barton, Kenneth A
; TITLE OF INVENTION: Immobilized Proteins in Cotton Fiber
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles and Brady
; STREET: P.O. Box 2113
; CITY: Madison
; STATE: WI
; COUNTRY: USA
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/217,327
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/812,233
; FILING DATE: 19-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 1122990831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 306 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-217-327-6
;
; Query Match 25.1%; Score 175; DB 1; Length 306;
; Best Local Similarity 34.6%; Pred. No. 6.5e-09;
; Matches 46; Conservative 4; Mismatches 41; Indels 42; Gaps 9
;
; QY 8 PPHGHHP-----HGPPPHGHHPHGGPPP-HGHPP-----HGPPPHHPHGGPPPHGH----- 52
;

```


Query Match	23.0%;	Score 160.5;	DB 1;	Length 78;
Best Local Similarity	46.8%;	Pred. No. 3.7e-08;		
Matches 36;	Conservative	3;	Mismatches 33;	Indels 5;
				Gaps 3;

```

RESULT 13
US-08-222-798A-5
; Sequence 5, Application US/08222798A
; Patent No. 5804553
; GENERAL INFORMATION:
; APPLICANT: KOKRYAKOV, VLADMIR N.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: LEHRER, ROBERT I.
; TITLE OF INVENTION: PROPHENINS-ANTIBIOTIC PEPTIDES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W.
; CITY: Washington, D.C.
; STATE: CA
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/222,798A
; FILING DATE: 05-APR-1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:

```

Qy	67	pprhghrrhghrpprhghr	83
Db	63	ppp----pffrrppffgpp	76

RESULT 12
US-08-487-359-5
; Sequence 5, Application US/08487359
; Patent No. 5633229
; GENERAL INFORMATION:
; APPLICANT: KOKRYAKOV, VLADMIR N
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: LEHRER, ROBERT I.
; TITLE OF INVENTION: PROPHENINS-
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

```
Run on: September 23, 2004, 22:55:11 ; Search time 68.9558 Seconds
        (without alignments)
        470.989 Million cell updates/sec
```

Title: US-10-074-225A-6
Perfect score: 697
Sequence: 1 SVNIIHRPPPHGHHPHGPPP.....HPPHGHGHHDHGPCDPPSHK 101

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1349238 seqs, 321558718 residues
Total number of hits satisfying chosen parameters: 1349238

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```

Database : Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	697	100.0		101	14	US-10-074-225A-6	Sequence 6, Appli
2	697	100.0		526	14	US-10-074-225A-3	Sequence 3, Appli
3	314	45.1		525	10	US-09-919-039-62	Sequence 62, Appl
4	314	45.1		525	14	US-10-074-225A-1	Sequence 1, Appli
5	280.5	40.2		148	14	US-10-074-225A-5	Sequence 5, Appli
6	248.5	35.7		235	16	US-10-437-963-171176	Sequence 171176,
7	246	35.3		274	9	US-09-850-887-4	Sequence 4, Appli
8	233.5	33.5		241	12	US-10-424-599-223180	Sequence 223180,
9	233.5	33.5		309	16	US-10-751-014-2	Sequence 2, Appli
10	233.5	33.5		309	16	US-10-751-014-5	Sequence 5, Appli
11	232.5	33.4		342	12	US-10-424-599-281824	Sequence 281824,
12	231	33.1		173	16	US-10-408-765A-69	Sequence 69, Appl
13	228.5	32.8		93	11	US-09-997-003-31	Sequence 31, Appl
14	228.5	32.8		93	11	US-09-997-003-44	Sequence 44, Appl
15	227.5	32.6		107	12	US-10-424-599-234553	Sequence 234553,

16	225	32.3	204	12	US-10-424-599-221495	Sequence 221495,
17	223	32.0	399	12	US-10-424-599-191976	Sequence 191976,
18	223	32.0	1130	14	US-10-032-585-7758	Sequence 7758, Ap
19	221.5	31.8	183	12	US-10-424-599-218680	Sequence 218680,
20	221.5	31.8	208	16	US-10-437-963-103915	Sequence 103915,
21	219	31.4	296	12	US-10-424-599-224246	Sequence 224246,
22	218.5	31.3	147	12	US-10-424-599-154677	Sequence 154677,
23	217.5	31.2	134	12	US-10-424-599-208512	Sequence 208512,
24	215.5	30.9	104	15	US-10-437-708-235	Sequence 235, App
25	215	30.8	179	12	US-10-424-599-218681	Sequence 218681,
26	213.5	30.6	449	15	US-10-369-493-12408	Sequence 12408, A
27	213	30.6	184	12	US-10-424-599-281818	Sequence 281818,
28	212	30.4	163	12	US-10-424-599-281820	Sequence 281820,
29	212	30.4	178	12	US-10-424-599-264218	Sequence 264218,
30	211.5	30.3	241	12	US-10-424-599-163337	Sequence 163337,
31	211	30.3	518	15	US-10-369-493-3240	Sequence 3240, Ap
32	210	30.1	126	12	US-10-424-599-281122	Sequence 281122,
33	208	29.8	865	9	US-09-957-995A-19	Sequence 19, Appl
34	207.5	29.8	105	12	US-10-424-599-171754	Sequence 171754,
35	207.5	29.8	189	16	US-10-767-701-36066	Sequence 36066, A
36	207	29.7	267	16	US-10-437-963-130670	Sequence 130670,
37	205	29.4	153	12	US-10-424-599-159963	Sequence 159963,
38	204	29.3	154	12	US-10-424-599-177249	Sequence 177249,
39	203	29.1	342	16	US-10-437-963-112806	Sequence 112806,
40	202.5	29.1	192	16	US-10-437-963-143333	Sequence 143333,
41	202	29.0	212	12	US-10-424-599-211931	Sequence 211931,
42	202	29.0	431	16	US-10-437-963-204963	Sequence 204963,
43	202	29.0	644	16	US-10-437-963-116147	Sequence 116147,
44	200	28.7	210	12	US-10-424-599-215748	Sequence 215748,
45	199	28.6	399	12	US-10-403-571-102	Sequence 102, App

ALIGNMENTS

RESULT 1

US-10-074-225A-6
; Sequence 6, Application US/10074225A
; Publication No. US20030082740A1
; GENERAL INFORMATION:
; APPLICANT: DONATE, Fernando
; APPLICANT: PLUNKETT, Marian L
; APPLICANT: HARRIS, Scott
; APPLICANT: MAZAR, Andrew P
; TITLE OF INVENTION: HISTIDINE PROLINE RICH GLYCOPROTEIN (HPRG) AS AN ANTI-ANGIOGENIC
; TITLE OF INVENTION: ANTI-TUMOR AGENT
; FILE REFERENCE: 38342-178463
; CURRENT APPLICATION NUMBER: US/10/074,225A
; CURRENT FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: US 60/268,370
; PRIOR FILING DATE: 2001-02-14
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Lepus americanus
; US-10-074-225A-6

	Query Match	100.0%; Score 697; DB 14;	Length 101;
	Best Local Similarity	100.0%; Pred. No: 1.6e-42;	
Matches	101; Conservative	0; Mismatches	0; Indels 0; Gaps 0;
Oy	1	SVNIHRPPPHGHHPHGGPPP 	PHGHHDPGPPPHGHPHPHGPBRP
Dd	1	SVNIIHRPPPHGHHPHGGPPP 	HGHPHPHGPBRP
Oy	61	GHPHGGPPP 	PHGHPHGHGFHDHGCDPSHK
Dd	61	GHPHGGPPP 	PHGHPHGHGFHDHGCDPSHK

RESULT 2

US-10-074-225A-3
; Sequence 3, Application US/10074225A
; Publication No. US20030082740A1
; GENERAL INFORMATION:
; APPLICANT: DONATE, Fernando
; APPLICANT: PLUNKETT, Marian L
; APPLICANT: HARRIS, Scott
; APPLICANT: MAZAR, Andrew P
; TITLE OF INVENTION: HISTIDINE PROLINE RICH GLYCOPROTEIN (HPRG) AS AN ANTI-ANGIOGENIC
; TITLE OF INVENTION: ANTI-TUMOR AGENT
; FILE REFERENCE: 38342-178463
; CURRENT APPLICATION NUMBER: US/10/074,225A
; CURRENT FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: US 60/268,370
; PRIOR FILING DATE: 2001-02-14
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 526
; TYPE: PRT
; ORGANISM: Lepus americanus
US-10-074-225A-3

Query Match 100.0%; Score 697; DB 14; Length 526;
Best Local Similarity 100.0%; Pred. No. 5.6e-42;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVNIHRPPPHGHHPPHGGPPPHGHHPPHGGPPPHGHHPPHGGPPPHGHHPPHGGPPPH 60
Db 321 SVNIHRPPPHGHHPPHGGPPPHGHHPPHGGPPPHGHHPPHGGPPPHGHHPPHGGPPPH 380
QY 61 GHPPHGGPPHGGPPHGGPPHGGPPHGGPPHGGPPHGGPPHGGPPHGGPPHGGPPH 101
Db 381 GHPPHGGPPHGGPPHGGPPHGGPPHGGPPHGGPPHGGPPHGGPPHGGPPHGGPPH 421

RESULT 3
US-09-919-039-62
; Sequence 62, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 62
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030108871A1 085596CD1
US-09-919-039-62

Query Match 45.1%; Score 314; DB 10; Length 525;
Best Local Similarity 41.8%; Pred. No. 5.9e-15;
Matches 64; Conservative 2; Mismatches 23; Indels 64; Gaps 8;

QY 7 RPP--PHG---HHPHGPPPHGHHPPHGGPP---HGHP-PHGPPP----- 40
Db 275 KPFFKPHGSRDHHHPK-----HEHGPPPPPPDERDHSHPPLPQGPPPLLPMSCSSCQH 329
QY 41 -----RHP-----PHGPPPHGHP-----HGPPPHGHPHP 67
Db 330 ATFTNGAQRSHNNSSDLHPKHSHEQHPHGHHPHHAHPHEHDTHRQHPHGHHPHGH 389
QY 68 PPHGHPHGGPPPHGHHPPHGGPPHGGPPHGGPPHGGPPHGGPPHGGPPH 100

Db 390 HPHGHHPPHGHHPHGHHPHCHDFQDYGCDPPPH 422

RESULT 4
US-10-074-225A-1
; Sequence 1, Application US/10074225A
; Publication No. US20030082740A1
; GENERAL INFORMATION:
; APPLICANT: DONATE, Fernando
; APPLICANT: PLUNKETT, Marian L
; APPLICANT: HARRIS, Scott
; APPLICANT: MAZAR, Andrew P
; TITLE OF INVENTION: HISTIDINE PROLINE RICH GLYCOPROTEIN (HPRG) AS AN ANTI-ANGIOGENIC
; TITLE OF INVENTION: ANTI-TUMOR AGENT
; FILE REFERENCE: 38342-178463
; CURRENT APPLICATION NUMBER: US/10/074,225A
; CURRENT FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: US 60/268,370
; PRIOR FILING DATE: 2001-02-14
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-074-225A-1

Query Match 45.1%; Score 314; DB 14; Length 525;
Best Local Similarity 41.8%; Pred. No. 5.9e-15;
Matches 64; Conservative 2; Mismatches 23; Indels 64; Gaps 8;

QY 7 RPP--PHG---HHPHGPPPHGHHPPHGGPP---HGHP-PHGPPP----- 40
Db 275 KPFFKPHGSRDHHHPK-----HEHGPPPPPPDERDHSHPPLPQGPPPLLPMSCSSCQH 329
QY 41 -----RHP-----PHGPPPHGHP-----HGPPPHGHPHP 67
Db 330 ATFTNGAQRSHNNSSDLHPKHSHEQHPHGHHPHHAHPHEHDTHRQHPHGHHPHGH 389
QY 68 PPHGHPHGGPPPHGHHPPHGGPPHGGPPHGGPPHGGPPHGGPPHGGPPH 100
Db 390 HPHGHHPPHGHHPHGHHPHCHDFQDYGCDPPPH 422

RESULT 5
US-10-074-225A-5
; Sequence 5, Application US/10074225A
; Publication No. US20030082740A1
; GENERAL INFORMATION:
; APPLICANT: DONATE, Fernando
; APPLICANT: PLUNKETT, Marian L
; APPLICANT: HARRIS, Scott
; APPLICANT: MAZAR, Andrew P
; TITLE OF INVENTION: HISTIDINE PROLINE RICH GLYCOPROTEIN (HPRG) AS AN ANTI-ANGIOGENIC
; TITLE OF INVENTION: ANTI-TUMOR AGENT
; FILE REFERENCE: 38342-178463
; CURRENT APPLICATION NUMBER: US/10/074,225A
; CURRENT FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: US 60/268,370
; PRIOR FILING DATE: 2001-02-14
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 148
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-074-225A-5

Query Match 40.2%; Score 280.5; DB 14; Length 148;
Best Local Similarity 51.6%; Pred. No. 5.3e-13;
Matches 47; Conservative 1; Mismatches 24; Indels 19; Gaps 2;

QY 10 PHGHHPPHGGPPPHGHHPPHGGPPPHGHHPPHGGPPPHGHHPPHGGPPPHGHHPPHGGPPPH 69

Db 27 KPPPEYQPPHEKPPH---ENPPPEHQPPHEKPPHEHQPPHEKPPPEYEPPEKPPPEYQ 82

QY 64 PH-GPPPHGHPPH-GPPPHGHPPHGHGHGFDHGPC--DPPSHK 101
 Db 83 PHEKPPPEYQPPHEKPPPEYQPPHEKPPPEYQPPHEKPPPEYQPPHEKPPPEHQ 124

RESULT 9
 US-10-751-014-2
 ; Sequence 2, Application US/10751014
 ; Publication No. US20040172676A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Franssen, Henk J
 ; APPLICANT: Bisseling, Anton H
 ; TITLE OF INVENTION: ENOD2 Gene Regulatory Region
 ; FILE REFERENCE: MPS 4-87FD2
 ; CURRENT APPLICATION NUMBER: US/10/751,014
 ; CURRENT FILING DATE: 2004-01-02
 ; PRIOR APPLICATION NUMBER: US/09/564,142A
 ; PRIOR FILING DATE: 2000-05-03
 ; PRIOR APPLICATION NUMBER: 08/859,555
 ; PRIOR FILING DATE: 1997-05-20
 ; NUMBER OF SEQ ID NOS: 6
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 2
 ; LENGTH: 309
 ; TYPE: PRT
 ; ORGANISM: Glycine max
 US-10-751-014-2

Query Match 33.5%; Score 233.5; DB 16; Length 309;
 Best Local Similarity 48.0%; Pred. No. 1.9e-09;
 Matches 49; Conservative 3; Mismatches 39; Indels 11; Gaps 7;

QY 7 RPPPHGHPPH-GPPPHGHPPHGHGHGFDHGPC--DPPSHK 101
 Db 95 KPPPEYQPPHEKPPH---ENPPPEHQPPHEKPPPEYEPPEKPPPEYQ 150
 QY 64 PH-GPPPHGHPPH-GPPPHGHPPHGHGHGFDHGPC--DPPSHK 101
 Db 151 PHEKPPPEYQPPHEKPPPEYQPPHEKPPPEYQPPHEKPPPEHQ 192

RESULT 10
 US-10-751-014-5
 ; Sequence 5, Application US/10751014
 ; Publication No. US20040172676A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Franssen, Henk J
 ; APPLICANT: Bisseling, Anton H
 ; TITLE OF INVENTION: ENOD2 Gene Regulatory Region
 ; FILE REFERENCE: MPS 4-87FD2
 ; CURRENT APPLICATION NUMBER: US/10/751,014
 ; CURRENT FILING DATE: 2004-01-02
 ; PRIOR APPLICATION NUMBER: US/09/564,142A
 ; PRIOR FILING DATE: 2000-05-03
 ; PRIOR APPLICATION NUMBER: 08/859,555
 ; PRIOR FILING DATE: 1997-05-20
 ; NUMBER OF SEQ ID NOS: 6
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 5
 ; LENGTH: 309
 ; TYPE: PRT
 ; ORGANISM: Glycine max
 US-10-751-014-5

Query Match 33.5%; Score 233.5; DB 16; Length 309;
 Best Local Similarity 48.0%; Pred. No. 1.9e-09;
 Matches 49; Conservative 3; Mismatches 39; Indels 11; Gaps 7;

QY 7 RPPPHGHPPH-GPPPHGHPPHGHGHGFDHGPC--DPPSHK 101
 Db 95 KPPPEYQPPHEKPPH---ENPPPEHQPPHEKPPPEYEPPEKPPPEYQ 150

QY 64 PH-GPPPHGHPPH-GPPPHGHPPHGHGHGFDHGPC--DPPSHK 101
 Db 151 PHEKPPPEYQPPHEKPPPEYQPPHEKPPPEYQPPHEKPPPEHQ 192

RESULT 11
 US-10-424-599-281824
 ; Sequence 281824, Application US/10424599
 ; Publication No. US20040031072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kovalic David K
 ; APPLICANT: Zhou Yihua
 ; APPLICANT: Cao Yongwei
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53223)B
 ; CURRENT APPLICATION NUMBER: US/10/424,599
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 285684
 ; SEQ ID NO 281824
 ; LENGTH: 342
 ; TYPE: PRT
 ; ORGANISM: Glycine max
 ; NAME/KEY: unsure
 ; LOCATION: (1)..(342)
 ; OTHER INFORMATION: unsure at all xaa locations
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT3847_96507C.1.pap
 US-10-424-599-281824

Query Match 33.4%; Score 232.5; DB 12; Length 342;
 Best Local Similarity 48.4%; Pred. No. 2.4e-09;
 Matches 61; Conservative 1; Mismatches 25; Indels 39; Gaps 16;

QY 8 RPPPHGHPPH-GPPPHGHPPHGHGHGFDHGPC--DPPSHK 101
 Db 179 PVPVHPYPH---PHPHPHPH-PHPHPHPYVYHSPPP--PPKPYVYHSPPPVPVHHHPYP 232
 QY 55 H-GPPPHGHPPH-GPPPHGHPPHGHGHGFDHGPC--DPPSHK 101
 Db 233 HPHPHPHPHPPPPKPYVYHSPPPVPVHPHPYH-PHPHPYHPHPHPHPYVYHS 291
 QY 96 DPPSHK 101
 Db 292 PVPVHPY 297

RESULT 12
 US-10-408-765A-69
 ; Sequence 69, Application US/10408765A
 ; Publication No. US20040101874A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Ghosh, Soumitra S.
 ; APPLICANT: Fahy, Eoin D.
 ; APPLICANT: Zhang, Bing
 ; APPLICANT: Gibson, Bradford W.
 ; APPLICANT: Taylor, Steven W.
 ; APPLICANT: Glenn, Gary M.
 ; APPLICANT: Warnock, Dale E.
 ; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
 ; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
 ; FILE REFERENCE: 660088.465
 ; CURRENT APPLICATION NUMBER: US/10/408,765A
 ; CURRENT FILING DATE: 2003-04-04
 ; NUMBER OF SEQ ID NOS: 3077
 ; SOFTWARE: FastSeq for Windows version 4.0
 ; SEQ ID NO 69
 ; LENGTH: 173
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens

Blank